МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ МІСЬКОГО ГОСПОДАРСТВА імені О. М. БЕКЕТОВА

О. О. ВОРОНКОВ

КОНСПЕКТ ЛЕКЦІЙ
з курсу

ОПТИМІЗАЦІЙНІ МЕТОДИ І МОДЕЛІ

(для студентів заочної форми навчання освітньо-кваліфікаційного рівня «бакалавр» напряму підготовки 6.030504 - Економіка підприємства)

Харків
ХНУМГ ім. О. М. Бекетова
2016

Воронков О. О. Конспект лекцій з курсу «Оптимізаційні методи і моделі» (для студентів заочної форми навчання освітньо-кваліфікаційного рівня «бакалавр» напряму підготовки 6.030504 - Економіка підприємства) / О. О. Воронков; Харків. нац. ун-т міськ. госп-ва ім. О. М. Бекетова. – Харків: ХНУМГ ім. О. М. Бекетова, 2016. – 110 с.

Автор: канд. екон. наук О. О. Воронков

Рецензент: канд. екон. наук, доц. Г. І. Базецька

Рекомендовано кафедрою економіки підприємств міського господарства, протокол № 1 від 27 серпня 2015 р.

ЗМІСТ

ВСТУП 5

ЗМ 1 Особливості і сфери застосування оптимізаційних методів і моделей в економіці. Класифікація. Лінійні оптимізаційні моделі 7

Тема 1 Предмет і особливості застосування оптимізаційних методів в економіці 7

1.1 Предмет дисципліни 7

1.2 Поняття про моделювання 9

1.3 Основні поняття оптимізаційних задач і моделей 11

Контрольні запитання 15

Тема 2 Лінійні оптимізаційні моделі 15

2.1 Загальна форма задачі лінійного програмування (ЗЗЛП) 16

2.2 Основні властивості ЗЗЛП та її перша геометрична інтерпретація 16

2.2.1 Основні поняття лінійної алгебри та опуклого аналізу, застосовувані в теорії оптимізації 16

2.2.2 Перша геометрична інтерпретація ЗЛП і графічний метод розв'язання 18

2.2.3 Основні теореми лінійного програмування 20

2.3 Канонічна форма задачі лінійного програмування (КЗЛП) 21

2.3.1 Базисні розв'язки канонічної задачі лінійного програмування 22

2.3.2 Властивості базисних планів задачі лінійного програмування 23

2.4 Симплекс-метод 24

2.4.1 Загальна характеристика 24

2.4.2 Таблична реалізація симплекс-методу 27

2.4.3 Збіжність симплекс-методу. Виродженість у задачах ЛП 29

Контрольні запитання 31

Тема 3 Транспортна задача 32

3.1 Транспортна задача в матричній постановці та її властивості 32

3.2 Методи побудови опорного плану 33

3.3 Метод потенціалів 36

3.4 Випадок виродження 40

3.5 Транспортна задача за критерієм часу 41

Контрольні запитання 43

ЗМ 2 Економічна інтерпретація і аналіз оптимальних планів лінійних оптимізаційних моделей 44

Тема 4 Теорія подвійності та двоїсті оцінки в аналізі розв'язків лінійних оптимізаційних моделей 44

4.1 Пряма та двоїста задачі як пара сполучених задач 44

4.2 Основні теореми подвійності, їхній економічний зміст 46

4.3 Двоїсті оцінки й дефіцитність ресурсів 49

Контрольні запитання 49

Тема 5. Аналіз лінійних моделей оптимізаційних задач 50
5.1 Аналіз розв’язків лінійних оптимізаційних моделей 50
5.2 Аналіз параметричної стійкості розв’язків ЗЛП 54
5.3 Оцінка рентабельності виробленої продукції 56
5.4 Аналіз обмежень дефіцитних і недефіцитних ресурсів 57
Контрольні запитання 57
ЗМ 3. Методи розв’язання нелінійних оптимізаційних задач 58
Тема 6 Цілочислові та дрібно-лінійні задачі лінійного програмування 58
6.1 Типи задач дискретного програмування 58
6.2 Метод Гоморі 60
6.3 Метод віток і границь 63
6.4 Дрібно-лінійне програмування 65
Контрольні запитання 66
Тема 7 Нелінійні оптимізаційні моделі 67
7.1 Постановка задачі нелінійного програмування (ЗНП) 67
7.2 Класичний метод оптимізації з використанням множників Лагранжа 69
7.3 Опукле програмування 70
7.4 Необхідні й достатні умови існування сідлової точки. Теорема Куна-Таккера 73
7.5 Деякі методи розв’язання задач НЛП 76
7.5.1 Градієнтні методи розв’язання задач безумовної оптимізації 76
7.5.2 Квадратичне програмування (КП) 78
Контрольні запитання 81
Тема 8 Динамічне програмування 82
8.1 Загальна схема методів динамічного програмування 82
8.2 Основні типи задач і моделі динамічного програмування 85
Контрольні запитання 90
Тема 9 Прийняття рішень в умовах невизначеності та ризику 91
9.1 Загальна математична постановка задачі стохастичного програмування (СП) 91
9.2 Класифікація задач стохастичного програмування. Основні методи розв'язання 92
9.3 Імітаційне моделювання 94
9.4 Прийняття рішень в умовах невизначеності й ризику 96
Контрольні запитання 100
Тема 10 Елементи теорії ігор 101
10.1 Основні поняття теорії ігор 101
10.2 Змішані стратегії. Основна теорема теорії ігор 104
10.3 Зведення гри двох гравців до задачі лінійного програмування 105
Контрольні запитання 107
Список джерел 109

ВСТУП

Дисципліна «Оптимізаційні методи і моделі» є нормативною дисципліною циклу природничо-наукової і загальноекономічної підготовки в навчальному плані за напрямом «Економіка і підприємництво» для кваліфікаційного рівня «Бакалавр». Обсяг курсу становить 144 академічних години або 4 кредити. При вивченні за заочною формою обсяг аудиторних занять становить 14 годин (8 годин лекцій і 6 годин практичних занять), на самостійну роботу студента припадає 130 годин. Програма курсу містить 3 змістових модулі: «Особливості і сфери застосування оптимізаційних методів і моделей в економіці. Класифікація оптимізаційних методів. Лінійні оптимізаційні моделі», «Економічна інтерпретація і аналіз оптимальних планів лінійних оптимізаційних моделей» і «Методи розв’язання нелінійних оптимізаційних задач», відповідно до яких виконується поточний контроль знань. Підсумковий контроль знань (іспит) проводиться в усній формі. У процесі вивчення курсу студенти мають виконати контрольну роботу.

Метою вивчення дисципліни «Оптимізаційні методи і моделі» є формування системи базових знань в області методології постановки задач, побудови оптимізаційних математичних моделей і методів їх розв'язання та аналізу.

В результаті вивчення курсу студенти повинні оволодіти прийомами побудови та аналізу оптимізаційних моделей, основними математичними поняттями і методами розв'язання оптимізаційних задач різної складності.

Дисципліна «Оптимізаційні методи і моделі» передбачає вивчення доволі широкого набору методичних прийомів, які сприяють ефективному дослідженню важливих проблем і задач, пов'язаних з управлінням організаційними системами (організаціями), що відповідає загальній меті підготовки фахівців, здатних комплексно використовувати на практиці спеціальні знання, математичні засоби дослідження складних процесів і систем, засоби обчислювальної техніки.

Управління організаційними системами можна розглядати як процес вироблення рішень з різноманітних питань. Процес прийняття рішень передбачає дослідження системи з метою отримання додаткової інформації та безпосередньо прийняття потрібного рішення на основі свого досвіду та інформації, що отримана у процесі дослідження.

Дослідження системи полягає у застосуванні наукових принципів і засобів до розв’язання задач, що виникають в процесі управління системою. Велику роль в одержанні інформації для прийняття рішень мають дослідження, що основані на математичному моделюванні. Найчастіше без належних кількісних характеристик, що отримані математичним шляхом, неможливо приймати ефективні рішення. Проте жодне дослідження не в змозі врахувати множину різноманітних факторів, які впливають на систему. Тому для прийняття рішення необхідні і мають істотну роль судження керівника, що основані на його досвіді, інтуїції, знаннях і творчих спроможностях. В реальному процесі прийняття рішень дослідження і неформальні судження взаємно пов’язані.

Розробкою наукових методів розв'язання організаційно-управлінських задач займається розділ математики «Математичне програмування». У математичному програмуванні задачу управління розглядають як математичну задачу. Проте на відміну від багатьох інших математичних задач вона припускає не один розв'язок, а множину різних розв'язків. Це пов'язано з тим, що в задачах управління є, як правило, багато способів організації будь-якого процесу, які призводять до поставленої мети. Якщо є множина розв'язків, то виникає додаткова задача - вибрати з цієї множини розв'язків такий, що з певної точки зору є найкращим. Можна навести багато прикладів таких задач. Наприклад, з одного міста в інше можна проїхати, користуючись різними видами транспорту: залізничним, повітряним, водним, автобусним, автомобільним. Додатковою задачею можна вважати вибір найбільш вигідного виду транспорту з погляду часу проїзду, або вартості, зручності та ін.

Використання математичних методів надає нові можливості, і фахівцеві необхідно вміти формулювати й розв'язувати задачі з оптимізації.

ЗМ 1 Особливості і сфери застосування оптимізаційних методів і моделей в економіці. Класифікація. Лінійні оптимізаційні моделі

ТЕМА 1
ПРЕДМЕТ І ОСОБЛИВОСТІ ЗАСТОСУВАННЯ ОПТИМІЗАЦІЙНИХ МЕТОДІВ В ЕКОНОМІЦІ

1.1 Предмет дисципліни

Предметом дисципліни «Оптимізаційні методи і моделі» є математичні методи кількісного обґрунтування рішень у задачах, пов'язаних з **управлінням організаційними системами**.

Нагадаємо, що **системою** називається сукупність взаємопов’язаних елементів, об'єднаних **загальною метою** функціонування. Як **організаційну систему** розуміють певну організацію, наприклад, виробниче підприємство. В загальному випадку природа таких систем може бути різною, зокрема, це можуть бути біологічні, соціологічні або інші процеси.

Системи, що зустрічаються на практиці, залежно від їхньої структури й характеру зв'язків поділяються на детерміновані та імовірнісні. **Детермінованою** називається система, закони руху якої точно відомі й майбутню поведінку якої можна передбачити. Для **імовірнісної** системи не можна зробити точного прогнозування її майбутньої поведінки. Прикладом детермінованої системи може служити годинниковий механізм. Системи статистичного контролю продукції, системи прибуття кораблів у морські порти або запас товарів на складі, що має велику кількість постачальників і споживачів, є імовірнісними системами.

Метою системи називається певний (заданий ззовні або встановлюваний самою системою) найкращий кінцевий стан (наприклад, параметри вихідних характеристик), тобто деяка підмножина значень функції системи.

Структура системи зумовлюється розташуванням і взаємозв'язками між складовими системи, які утворені для виконання системою своєї функції. Ці складові називаються **елементами** системи. Отже структура системи залежить від величини й складності системи. Величина системи визначається числом її елементів і кількістю зв'язків між ними, а складність - різноманіттям елементів, неоднорідністю їхніх властивостей та різною якістю зв'язків.

Великим і складним системам, таким як промислове підприємство, притаманні властивості **цілісності** та **емерджентності**.

Цілісність системи означає, що всієї її частини слугують спільній меті й сприяють формуванню найкращих (оптимальних) результатів у смислі прийнятого критерію ефективності.

Емерджентність (поява нового) означає, що великі й складні системи мають такі якості, що не властиві жодному з її елементів. Емерджентні або системні якості кардинально відрізняють системи від не систем. Інакше кажучи, об'єднання підсистем з різною природою і структурою (наприклад, економічної, технічної та соціальної) у складну систему зумовлює взаємний вплив підсистем одна на одну. Це створює нову системну якість, яка не властива ані одній з підсистем.

Визначимо зміст терміну «управління». У широкому значенні слова під управлінням розуміють організаційну діяльність, що здійснює функції керівництва чужою роботою, спрямованою на досягнення певних цілей. Процес управління полягає в прийнятті рішень про найбільш доцільні дії в тій або іншій сформованій ситуації. Людина, що здійснює управління, приймає рішення, оцінюючи навколишню обстановку за допомогою інформації, одержуваної від своїх органів почуттів, вимірювальних приладів, інших осіб. У багатьох випадках цієї інформації недостатньо для однозначної оцінки обстановки. Тоді людина використовує свій досвід, свої знання, пам'ять, інтуїцію. Чудовою властивістю людини є здатність приймати рішення в умовах істотної невизначеності щодо навколишньої обстановки. Проте в умовах сучасних великих промислових підприємств знань та інтуїції навіть у досвідченого керівника недостає, щоб здійснювати ефективне управління. У результаті використовують математичні методи, зокрема оптимізаційні методи і моделі.

Управлінням називають цілеспрямований вплив на систему з метою отримати найкращий ефект. Можна сказати, що управління являє собою таку організацію того або іншого процесу, що забезпечує досягнення певних цілей.

Можна виділити чотири етапи, що характерні для будь-якого процесу управління: поява цілі, оцінка ситуації, ухвалення рішення та виконання ухваленого рішення. Оскільки етап появи цілі передує початку процесу управління, його можна виключити з розглядання. Врахуємо також, що при управлінні складними процесами оцінка ситуації проводиться на підставі зібраної та обробленої інформації. Таким чином, дістанемо наступні три етапи процесу управління: збір і обробка інформації з метою оцінки сформованої ситуації; ухвалення рішення про найбільш доцільні дії; виконання ухваленого рішення.

Управління завжди вимагає визначення певної ознаки (або критерію), що дозволяє оцінити ефективність прийнятого рішення. Критерій пояснює ступінь відповідності поведінки системи цілям управління. Найчастіше ефективність визначають як умову, за якої система збільшить доходи при постійних або менших витратах або матиме постійний прибуток при менших витратах та ін. Загалом, ефективність можна уявити як кращий (необхідний) результат функціонування системи за менший час, при витраті менших ресурсів у довгостроковій і контрольованій перспективі. Уявлення про результат (ціль) має принципові відзнаки. Наприклад, переваги виробничої системи з вкладення ресурсів у нове обладнання, навчання персоналу або в збільшення жалування й розвиток соціальної сфери, або підвищення виплат власникам мають різну ефективність з позицій персоналу, власника, менеджера та стороннього спостерігача. Таким чином, визначення ефективності пов'язано з системою виражених і прихованих уявлень про її сутність, цілі, місце розташування спостерігача. Тому потрібно відзначити виняткову складність задачі строгої формалізації критеріїв ефективності.

У багатьох випадках реалізація процесу управління вимагає витрати будь-яких ресурсів: витрат часу, витрати матеріалів, палива, електроенергії. Отже, при виборі способу управління потрібно говорити не тільки про те, чи досягається поставлена ціль, але й про те, які ресурси доведеться затратити для досягнення цієї цілі. Тоді завдання з управління полягає в тому, щоб з множини рішень, які забезпечують досягнення цілі, вибрати одно рішення, що вимагає найменшої витрати ресурсів. В інших випадках підставою для переваги одного способу управління перед іншим можуть служити інші вимоги: вартість обслуговування, надійність та ін.

Математичний вираз, що дає кількісну оцінку ступеня досягнення цілі управління, називається **критерієм ефективності** управління. Найбільш кращим або оптимальним способом управління буде такий, при якому критерій ефективності досягає **мінімального** або **максимального** значення. При виборі, наприклад, режиму польоту за критерій якості управління можна прийняти або вираз для кількості палива, що витрачається на одиницю шляху, або шлях, на який витрачається одиниця палива. Найбільш економічному, тобто оптимальному, режиму відповідатиме або мінімальне (у першому випадку), або максимальне (у другому випадку) значення критерію.

При розв'язанні задачі управління треба розуміти, що управління будь-якою системою завжди піддане різним обмеженням. Вони викликані обмеженістю ресурсів, використовуваних при управлінні, або інших величин, які в силу фізичних особливостей тієї або іншої системи не можуть перевищувати певних меж. Математично обмеження цього виду виражають зазвичай у вигляді систем алгебраїчних рівнянь або нерівностей, що зв'язують змінні, які описують стан системи. Тому особливе місце в методах оптимізації займає **критерій оптимізації з обмеженнями**. Смисл тут полягає в тому, що відшукується не будь-яке рішення x^*, що обертає критерій ефективності на мінімум, а таке, що задовольняє системі обмежень. Неважко переконатися, що значення умовного екстремуму не може бути меншим за значення абсолютного екстремуму (без обмежень). Задачі оптимізації при наявності обмежень призвели до перегляду класичних методів оптимізації і створенню нових методів, відомих як методи програмування.

1.2 Поняття про моделювання

Моделюванням називають побудову копії (моделі) будь-якого процесу або об'єкту. Завдання моделювання полягає у відтворенні явища, що подібне оригіналу, і його дослідженні.

Існує ряд визначень поняття «модель». Наведемо одне з них:

Модель - це такий матеріальний або уявлений об'єкт, який у процесі пізнання (вивчення) заміщує об'єкт-оригінал, зберігаючи деякі найважливіші для даного дослідження типові його риси.

Моделі призначені: для вивчення складних процесів, явищ, тощо. Не можна, наприклад, проводити експерименти з економікою країни, з минулим, із сонячною системою, планетами та ін.

Моделі використовують для виявлення найбільш істотних факторів, що формують ті або інші властивості об'єктів, а також дозволяють навчитися правильно управляти об'єктами.

Розрізняють моделі геометричні, фізичні та математичні. Геометричні моделі дають зовнішнє уявлення про об'єкт, що нас цікавить, і більшою частиною використовуються для демонстраційних цілей. Фізичні й математичні моделі призначені для визначення числових значень величин, що характеризують поведінку об'єкту шляхом модифікації відповідних величин в моделі. При фізичному моделюванні дослідження проводять на моделі, що відрізняється від оригіналу тільки розмірами.

Модель має бути подібною основному об'єкту не абсолютно, а тільки у ступені, необхідному для виявлення основних властивостей об'єкта. Використання моделювання як інструмента пізнання ґрунтується на теорії подібності, основні положення якої збігаються до наступного:

- наявність аналогії між двома процесами або явищами в одних властивостях дає основу для прийняття гіпотези про те, що якщо перший процес має властивість «А», то і другий процес має цю саму властивість;

- між аналогічними процесами або явищами можна встановити подібність. Подібними називають такі процеси і явища, між відповідними параметрами яких є пропорційність;

За наявності подібних процесів або явищ за поведінкою одного з них можна судити про поведінку іншого.

Об'єктивною ознакою аналогічності зовні різних процесів служить ідентичність математичних виразів, що описують їхню сутність.

Головна перевага математичного моделювання полягає у можливості дослідження явищ, які важко піддаються вивченню, на добре вивчених явищах. Цей вид моделювання використовує прості, дешеві та зручні елементи. Основні переваги засобів математичного моделювання дозволили їм посісти переважне місце при проведенні наукових досліджень. Таким чином, модель необхідна:

- для того, щоб зрозуміти, як влаштований конкретний об'єкт: яка його структура, основні властивості, закони розвитку і взаємодія з навколишнім світом;

- для того, щоб навчитися управляти об'єктом (або процесом) і визначити найкращі засоби управління при заданих меті і критеріях;

- для того, щоб прогнозувати прямі та побічні наслідки реалізації заданих засобів і форм впливу на об'єкт.

1.3 Основні поняття оптимізаційних задач і моделей

Економічні задачі, ціль яких полягає у знаходженні найкращого (оптимального) з погляду певного критерію або критеріїв варіанту використання наявних ресурсів (праці, капіталу та ін.), називають оптимізаційними.

Оптимізаційні задачі вирішують за допомогою оптимізаційних моделей за методами математичного програмування.

Структура оптимізаційної моделі включає цільову функцію, області припустимих рішень і системи обмежень, що визначають цю область. Цільова функція в самому загальному вигляді, у свою чергу, так само складається з трьох елементів: керованих змінних, некерованих змінних, форми функції (виду залежності між змінними).

Розглядаючи певну довільну систему рівнянь, яку складають m рівнянь з n невідомими, можна виділити три типи задач:

Якщо $m=n$ – має місце алгебраїчна задача, яка має одно розв’язок.

Якщо $m>n$ - має місце задача перевизначена, що, як правило, не має розв’язків.

Якщо $m<n$ - має місце задача невизначена, що має нескінченно багато розв’язків.

В економіці доводиться мати справу з задачами третього типу, тобто з такими, що мають нескінченно багато розв’язків.

Наведемо приклади оптимізаційних задач.

План постачання підприємств. Є ряд підприємств, що споживають певні види сировини, і є ряд сировинних баз, що можуть поставляти цю сировину підприємствам. Бази зв'язані з підприємствами певними шляхами сполучення (залізниці, водний, автомобільний транспорт) із своїми тарифами. Потрібно розробити такий план постачання підприємств сировиною (з якої бази, в якій кількості, яку сировину доставити), щоб потреби в сировині були цілком забезпечені при мінімальних витратах на перевезення.

Будівництво ділянки магістралі. Споруджується ділянка залізничної магістралі. В розпорядженні є певна кількість ресурсів: людей, будівельних машин і механізмів, ремонтних майстерень, вантажних автомобілів тощо. Потрібно так спланувати будівництво, тобто призначити черговість робіт, розподілити машини і людей по ділянках шляху, забезпечити ремонтні роботи, щоб воно було завершене у мінімально можливий термін.

Бібліотечне обслуговування. Велика бібліотека обслуговує запити, що надходять від абонентів. В фондах бібліотеки є книги, що користуються підвищеним попитом, книги, на які запити надходять рідко і, нарешті, книги, на які запити майже ніколи не надходять. Є ряд можливостей розподілу книг на стелажах і сховищах, а також з диспетчеризації запитань із зверненнями до інших бібліотек. Потрібно розробити таку систему бібліотечного обслуговування, при якій запити абонентів задовольнялися би у максимальному ступені.

Математичне формулювання оптимізаційної задачі полягає у формулюванні цілі управління, яку виражають через критерій ефективності, визначення певних обмежень, що уявляють собою систему алгебраїчних рівнянь або нерівностей, які виражають обмеженість ресурсів або інших величин, використовуваних при управлінні.

Рішення про спосіб управління, що задовольняє всім поставленим обмеженням і перетворює на мінімум (максимум) критерій ефективності, називається **оптимальним рішенням**.

Отже, сутність всіх оптимізаційних задач збігається до пошуку такого розв'язку $\overline{X}$, що перетворює на екстремум критерій ефективності, виражений як функція від елементів прийнятого розв'язку $\overline{X}$ й називаний **цільовою функцією** $F(x)$. Потрібно відзначити, що методи розв’язання оптимізаційних задач є не аналітичною, а алгоритмічною формою розв'язання задач, тобто дають не формулу, яка виражає остаточний результат, а вказують лише обчислювальну процедуру, що призводить до розв'язання задачі. Тому оптимізаційні методи стають ефективними головним чином при використанні обчислювальної техніки.

Відзначимо, що до оптимізаційних задач, як правило, незастосовні методи класичного аналізу для відшукання умовних екстремумів. Це зумовлене такими специфічними їхніми особливостями:

1) коли на елементи розв'язку $\overline{X}$ накладені обмеження, екстремум найскоріше досягається не в точках, де похідні дорівнюють нулю, а на границі області обмежень;
2) у практичних задачах число змінних і число обмежень настільки велике, що пошук екстремуму шляхом визначення похідних є не ефективним;
3) у багатьох задачах математичного програмування цільова функція не має похідних (наприклад, задана тільки для цілочислових значень аргументів).

У зв'язку з цим метою математичного програмування є створення аналітичних методів з визначення розв'язку або ефективних обчислювальних способів одержання наближеного розв'язку оптимізаційної задачі.

Математичне моделювання економічних процесів є, з одного боку, дуже важливим і складним, а з іншого боку – таким, що практично не піддається науковій формалізації. Неодноразові спроби, виділити загальні принципи створення математичних моделей призводили або до декларування рекомендацій самого загального характеру, які важко застосовувати для розв'язання конкретних проблем, або, навпаки, до появи рекомендацій, що можна застосувати у дійсності тільки до вузького кола задач. Тому доцільно вивчати техніку математичного моделювання на конкретних прикладах.

Спільним для оптимізаційних задач є проблема пошуку найбільшого або найменшого (**оптимального**) значення певної функції, що відбиває **ціль управління** системою, або, як ще кажуть, **цільової функції**. Пошук оптимального значення здійснюється на певній підмножині припустимих значень змінних, що описують стан цієї системи, який називається **множиною припустимих планів**.

Нехай на певній множині D визначено функцію $f(x)$. Нагадаємо, що точка x^*, що належить D ($x^* \in D$), називається **точкою глобального максимуму**, якщо для будь-якого $x \in D$ виконується нерівність $f(x) \le f(x^*)$. У цьому випадку значення $f(x^*)$ називається **глобальним максимумом функції**. Точка $\hat{x}$ називається **точкою локального максимуму**, якщо існує певне оточення цієї точки, у будь-якій точці якого значення функції менші, ніж в $\hat{x}$ ($f(x) \le f(\hat{x})$). Аналогічно визначаються **глобальний** і **локальний мінімуми**. Узагальнюючим поняттям для максимуму й мінімуму є такий термін, як **екстремум** (оптимум).

Необхідно відзначити, що далеко не завжди весь комплекс цілей і задач під час моделювання об'єкту можна виразити у формі певної цільової функції. Більше того, усвідомлення й осмислення цієї проблеми стало свого роду переломним етапом в історії розвитку даної науки. Це дало поштовх до розвитку нових напрямків, що пов'язані з методами **багатокритеріальної (або векторної) оптимізації**, коли критерієм оптимальності є вимога мінімізації або максимізації кількох скалярних функцій. Проте всі вони базуються на методах **однокритеріальної оптимізації**, без ясного розуміння яких неможлива робота з більш складним математичним апаратом.

Потужним інструментом вирішення подібних задач стали спеціальні методи пошуку екстремуму, що складають зміст математичного програмування. У цьому випадку поняття програмування вживається як планування (на відміну від програмування для ЕОМ).

Задачі математичного програмування мають велику розмаїтість. Їх математичне моделювання практично не піддається науковій формалізації через те, що принцип побудови математичної моделі істотно залежить від конкретної природи досліджуваної системи. Проте у цих задачах прийнято виділяти певну послідовність етапів дослідження, зокрема:

1. Постановка задачі;
2. Словесне формулювання задачі з визначенням цілі її розв'язання та факторів-обмежень, що впливають на нього (вербальна модель);
3. Формалізація задачі – побудова адекватної математичної моделі. На цьому етапі цільова функція $F(x)$ виражається як залежність від розв'язку $\overline{X}$, а обмеження записують у вигляді системи рівностей і нерівностей;
4. Розв'язання задачі на базі математичної моделі;
5. Перевірка отриманих результатів на їхню адекватність природі досліджуваної системи, можливе коректування первісної моделі;
6. Розробка рекомендацій на підставі отриманого розв'язку.

Введемо ряд визначень.

Розв'язком (або **планом**) називається певний вибір параметрів, що залежать від нас. Параметри, сукупність яких утворює розв'язок, називають **елементами розв'язку**. Як елементи розв'язку можуть бути числа, вектори, функції, фізичні ознаки та ін.

$$x = (x_1, x_2, \ldots, \ldots, x_j, \ldots, x_n).$$

Система обмежень за ресурсами формує множину припустимих розв'язків (планів) D. Той факт, що розв'язок x належить множині припустимих розв'язків D, записується в такий спосіб:

$$x \in D.$$

Оптимальним розв'язком або **оптимальним планом** називають таке рішення, що обертає цільову функцію $F(x)$ на максимум або мінімум.

Отже, у найбільш загальному вигляді оптимізаційна задача формулюється в такий спосіб:

При заданих обмеженнях знайти таке рішення $x = x^*$, що обертає цільову функцію $F(x)$ на максимум або мінімум.

$$F^* = \underset{x \in D}{extr}\{F(x, \alpha)\},$$

де α - система обмежень задачі.

Залежно від вигляду цільової функції $F(x)$ і системи обмежень α виділяють наступні методи розв'язання оптимізаційних задач.

Лінійне програмування. Застосовується, якщо в моделі цільова функція $F(x)$ є лінійною, а множина D, на якій шукають її екстремум, задається системою лінійних рівнянь і нерівностей. У лінійному програмуванні існує клас задач, структура яких дозволяє створити спеціальні методи розв'язання, що вигідно відрізняються від методів розв'язання задач загального характеру, зокрема - транспортна задача.

Нелінійне програмування. Тут є нелінійними цільова функція й обмеження. У нелінійному програмуванні виділяють такі класи задач:

- **опукле програмування** – коли цільова функція є опуклою (якщо розглядають задачу її мінімізації) і опуклою є множина, на якій вирішується екстремальна задача;

- **квадратичне програмування** – коли цільова функція є квадратичною, а обмеження – лінійні рівності або нерівності.

Дискретне програмування. Даний метод використовують, коли на елементи рішення x накладено вимогу дискретності, наприклад, цілочисленості. Така вимога істотно ускладнює розв'язання задачі, тому що застосування стандартних прийомів (вирішити задачу як аналогову, а потім округлити результат до цілого значення) неможливо.

Динамічне програмування. Це метод, що дозволяє шляхом покрокової оптимізації певних проміжних цільових функцій отримати загальний результуючий оптимум. У задачах динамічного програмування цільова функція $F(x)$ є адитивною або мультиплікативною функцією змінних x.

Стохастичне програмування. Даний вид програмування використовується, коли параметри умов або елементи розв'язку є випадковими величинами, що зумовлене невизначеністю, яка породжує ризикованість прийнятих рішень. У стохастичному програмуванні труднощі виникають не тільки під час розробки методів розв'язання задач, а й під час їхньої постановки.

Евристичне програмування. Застосовують для розв'язання задач, у яких точний оптимум знайти алгоритмічним шляхом неможливо через величезне число варіантів. У такому випадку відшукують не оптимальний, а досить гарний з погляду практики розв'язок.

Контрольні запитання

1. Охарактеризуйте предмет і зміст науки «Оптимізаційні методи і моделі».
2. Охарактеризуйте особливості оптимізаційних задач.
3. Які загальні етапи розв'язання оптимізаційних задач прийнято виділяти?
4. Чому до оптимізаційних задач не застосовують класичні методи пошуку умовного екстремуму функції?
5. Що являє собою цільова функція оптимізаційної задачі? Яке її призначення?
6. Дайте визначення понять: план, припустимий план, оптимальний план, розв'язок задачі.
7. На чому базується класифікація моделей і методів розв'язання оптимізаційних задач? Які класи моделей і методів виділяють у математичному програмуванні?
8. Що являє собою множина можливих розв'язків оптимізаційної задачі?
9. Поясніть, яка область можливих розв'язків оптимізаційної задачі називається областю припустимих планів.

ТЕМА 2
ЛІНІЙНІ ОПТИМІЗАЦІЙНІ МОДЕЛІ

Серед оптимізаційних задач найпростішими й щонайкраще розробленими є лінійні оптимізаційні моделі, для розв’язання яких використовують методи лінійного програмування. Характерним для них є те, що:

- цільова функція $F(x)$ лінійно залежить від елементів розв'язку $x_1,x_2,\ldots,x_n$. Її називають **лінійною формою** й позначають L;
- обмеження, що накладають на елементи розв'язку, мають вигляд лінійних рівностей і нерівностей щодо $x_1, x_2,\ldots, \ldots,x_j, \ldots,x_n$.

Такі задачі часто зустрічаються на практиці. До задач лінійного програмування належить, зокрема, найпростіша задача виробничого планування (задача про оптимальне використання ресурсів). Також до задач лінійного програмування належать задачі про використання інвестицій, про мінімізацію витрат, транспортна задача, задача про складання раціону та ін.

Математичну модель задачі лінійного програмування завжди записують у двох формах – у **загальній формі** (ЗЗЛП) і **канонічній формі** (КЗЛП).

2.1 Загальна форма задачі лінійного програмування (ЗЗЛП)

У загальному вигляді задачу лінійного програмування (ЗЛП) формулюють в такий спосіб:

Знайти найбільше або найменше значення лінійної функції

$$L = \sum_{j=1}^{n} c_j x_j \to extr \, . \tag{2.1}$$

на певній множині D, де $x \in D$ задовольняє системі обмежень

$$\begin{gathered} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \le b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \le b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{k1}x_1 + a_{k2}x_2 + \ldots + a_{kn}x_n \le b_k, \\ a_{(k+1)1}x_1 + a_{(k+1)2}x_2 + \ldots + a_{(k+1)n}x_n = b_{(k+1)}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m, \\ x_j \ge 0, \quad j = \overline{1,n} \, . \end{gathered} \tag{2.2}$$

Відзначимо, що в системі перші k обмежень є нерівностями, а наступні m-k - рівняннями. Цього можна домогтися простим переупорядкуванням виразів.

Щодо направлення знака нерівності будемо вважати, що ліва частина менша або дорівнює правій частині. Домогтися цього можна, помноживши на (-1) обидві частини нерівностей із протилежним знаком.

Вибір типу шуканого екстремуму цільової функції також не принциповий, оскільки задача пошуку максимуму функції $L=\Sigma c_j x_j$ є еквівалентною задачі пошуку мінімуму функції $-L=\Sigma(-c_j x_j)$.

Задачу лінійного програмування, записану в такій формі, називають загальною задачею лінійного програмування (ЗЗЛП).

2.2 Основні властивості ЗЗЛП та її перша геометрична інтерпретація

2.2.1 Основні поняття лінійної алгебри та опуклого аналізу, застосовувані в теорії оптимізації

Коротко нагадаємо деякі фундаментальні визначення й теореми лінійної алгебри та опуклого аналізу, які широко застосовуються при розв'язанні задач як лінійного, так і нелінійного програмування.

Фундаментальним поняттям лінійної алгебри є лінійний (речовинний) простір. Під ним мають на увазі множину певних елементів (що називаються векторами або точками), для яких задані операції додавання й множення на речовинне число (скаляр), причому елементи, що є результатом виконання операцій, також відповідно до визначення мають належати вихідному простору.

Окремими випадками лінійних просторів є речовинна пряма, площина, геометричний тривимірний простір.

Вектор $\lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_m a_m$ називається **лінійною комбінацією** векторів $a_1, a_2, \ldots, a_m$ з коефіцієнтами $\lambda_1, \lambda_2, \ldots, \lambda_m$.

Система векторів лінійного простору $a_1, a_2, \ldots, a_m$ називається **лінійно залежною**, якщо існують такі числа $\lambda_1, \lambda_2, \ldots, \lambda_m$, які не дорівнюють одночасно нулю, що їхня лінійна комбінація $\lambda_1 a_1, \lambda_2 a_2, \ldots, \lambda_m a_m$ дорівнює нульовому вектору (вектору, усі компоненти якого дорівнюють нулю). У протилежному випадку систему $a_1, a_2, \ldots, a_m$ називають **лінійно незалежною**, тобто лінійна комбінація цих векторів може дорівнювати нульовому вектору тільки при нульових коефіцієнтах $\lambda_1, \lambda_2, \ldots, \lambda_m$.

Максимально можлива кількість векторів, які можуть утворювати лінійно незалежну систему в певному лінійному просторі, називають **розмірністю простору**, а будь-яку систему лінійно незалежних векторів у кількості, що дорівнює розмірності, - **базисом простору**.

Лінійний простір позначають R^n, де n - його розмірність.

Будь-яка підмножина даного лінійного простору, що саме має властивості лінійного простору, називається лінійним підпростором. Множина H, одержувана зсувом певного лінійного підпростору $L \in R^n$ на вектор $a \in R^n$: $H = L + a$, називається афінною множиною (простором). Якщо фундаментальною властивістю будь-якого лінійного простору або підпростору є приналежність йому нульового вектору, то для афінної множини це не завжди так. На площині прикладом підпростору є пряма, що проходить через початок координат, а афінної множини - будь-яка пряма на площині. Характерною властивістю афінної множини є належність їй будь-якої прямої, що з'єднує дві будь-які її точки. Розмірність афінної множини збігається з розмірністю того лінійного підпростору, зсувом якого вона отримана.

Якщо розглядається певний лінійний простір R^n, то приналежні йому афінні множини розмірності 1 називаються прямими, а розмірності (n-1) - гіперплощинами. Так, звичайна площина є гіперплощиною для тривимірного геометричного простору R^3, а пряма - гіперплощиною для площини R^2. Усяка гіперплощина поділяє лінійний простір на два півпростори.

Множина V векторів (точок) лінійного простору R^n називається **опуклою**, якщо вона містить відрізок прямої, що з'єднує дві будь-які його точки, або, інакше кажучи, з того, що $a \in V$ і $b \in V$, випливає, що $x = (1-\lambda)a + \lambda b \in V$, де $0 \le \lambda \le 1$.

Лінійна комбінація $\sum_{i=1}^{m} \lambda_i a_i$ векторів $a_1, a_2, \ldots, a_m$ називається **опуклою**, якщо $\lambda_i \ge 0$, $i = \overline{1, m}$ і $\sum_{i=1}^{m} \lambda_i = 1$.

Множину, що містить всі можливі опуклі комбінації точок певної множини M, називають опуклою оболонкою даної множини. Можна показати, що опукла оболонка множини M є найменшою опуклою множиною, що містить M.

Опукла оболонка кінцевої множини точок називається **опуклим багатогранником**, а непусте перетинання кінцевої кількості замкнутих

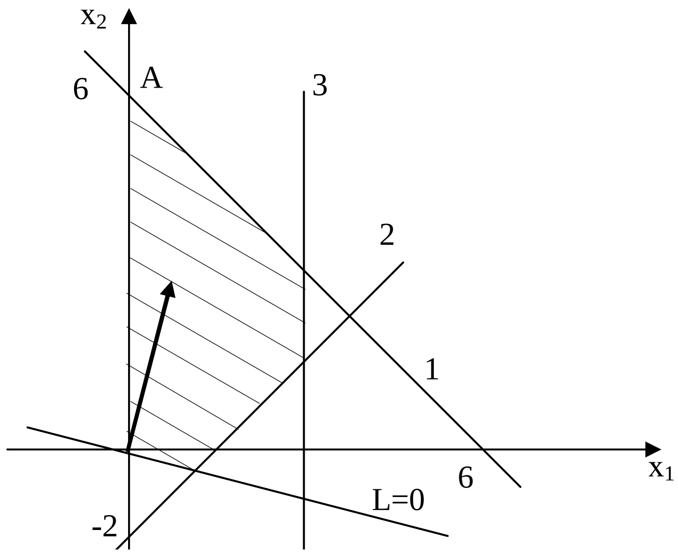

Рисунок 2.1 – Геометрична інтерпретація ЗЗЛП

півпросторів - **багатогранною опуклою множиною**. На відміну від опуклого багатогранника остання може бути необмеженою.

Точка v опуклої множини V називається його кутовою (крайньою) точкою, якщо вона не є внутрішньою точкою ні для якого відрізка, кінці якого належать множині V. Кутові точки опуклого багатогранника є його вершинами, а сам він - опуклою оболонкою своїх вершин.

2.2.2 Перша геометрична інтерпретація ЗЛП і графічний метод розв'язання

У тому випадку, коли ЗЗЛП містить дві змінні x_1 і x_2, її можна зобразити на координатній площині та одержати розв'язок графічним методом. Графічне розв'язання ЗЗЛП носить ілюстративний характер, але основний зміст і термінологія розповсюджуються на задачі великої розмірності.

Розглянемо приклад. Нехай цільова функція представлена виразом

$$L = x_1 + 3x_2 \to \max,$$

а обмеження задані системою нерівностей:

$$\begin{aligned} x_1 + x_2 &\le 6, \\ x_1 - x_2 &\le 2, \\ x_1 \quad &\le 3, \\ x_1 \ge 0,\ &x_2 \ge 0. \end{aligned}$$

Будемо зображувати пару значень x_1 і x_2 точкою на координатній площині $x_1 0 x_2$ з координатами (x_1, x_2), що показано на рисунку 2.1.

Кожна нерівність визначає певну напівплощину.

Перетинання трьох напівплощин являє собою множину припустимих планів D, тому що кожна точка його множини належить одночасно кожній із трьох напівплощин, а отже задовольняє обмеженням ЗЗЛП. Помітимо, що припустимих розв'язків - нескінченна множина.

Для визначення оптимального плану задачі, тобто такого розв'язку (x_1, x_2), що обертає цільову функцію на максимум, скористаємося визначеннями:

- градієнтом функції $f(x)$ називається вектор

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \ldots, \frac{\partial f(x)}{\partial x_n} \right),$$

який вказує напрям найбільш швидкого зростання функції $f(x)$;

- лінією рівня функції $f(x)$ називається множина точок з області її визначення, у яких функція приймає те саме фіксоване значення.

У нашому прикладі $\nabla L = (1, 3)$. Лінії рівня L перпендикулярні напрямку градієнта. Побудуємо опорну пряму L=0, що проходить через початок координат, і будемо переміщувати її в напрямку ∇L. Очевидно, що переміщати лінію рівня в напрямку зростання цільової функції має сенс тільки в межах області припустимих розв'язків. Точкою, у якій цільова функція дістане максимальне значення, у нашому прикладі є точка A з координатами (0, 6). Таким чином, отриманий оптимальний план задачі

$$x^* = (0, 6),$$

при якому цільова функція приймає максимальне значення

$$L = 1*0 + 3*6 = 18.$$

Теоретично можливі також такі окремі випадки розв'язку ЗЗЛП:

- цільова функція L не обмежена зверху, тобто не має максимуму (рис. 2.2);
- коли лінія рівня збігається із гранню області припустимих розв'язків (рис. 2.3). У цьому випадку всі точки, що лежать на грані множини D, є оптимальними планами й кажуть, що має місце **альтернативний оптимум**.

У розглянутих ілюстраціях припустимі плани ЗЗЛП мають вигляд опуклої багатогранної множини. Таке подання множини припустимих планів називається першою геометричною інтерпретацією ЗЗЛП.

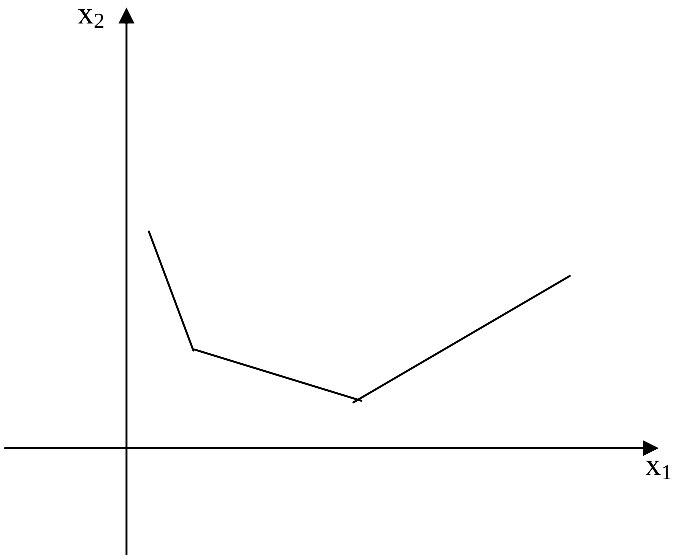

Рисунок 2.2 - Цільова функція L не обмежена зверху

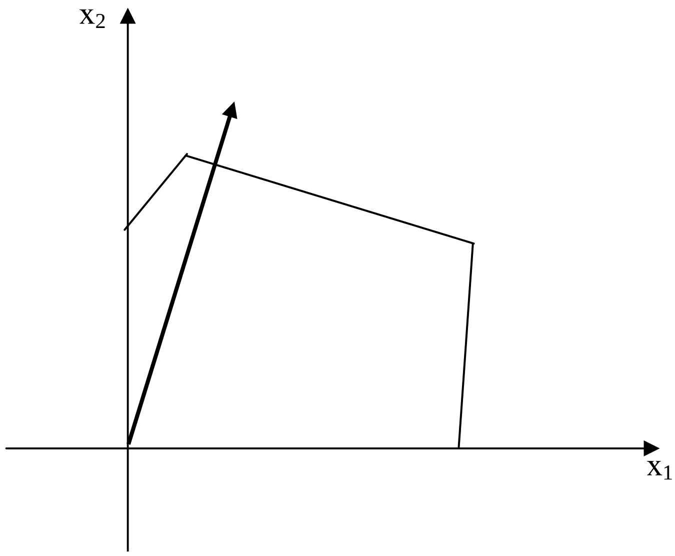

Рисунок 2.3 – Альтернативний оптимум

2.2.3 Основні теореми лінійного програмування

Розглянемо деякі теореми, що відбивають фундаментальні властивості задач лінійного програмування які полягають в основі методів їхнього розв'язання. Вони узагальнюють на випадок задач із довільною кількістю змінних ті властивості, які ми спостерігали у двовимірному випадку.

Теорема 2.1 Якщо цільова функція *L* приймає максимальне значення в певній точці множини припустимих планів *D*, то вона приймає це саме значення й у певній кутовій точці даної множини.

Доказ.

Щоб не ускладнювати виклад, обмежимося тим випадком, коли множина *D* обмежена, і, отже, є опуклим багатогранником. Для доказу скористаємося наступною відомою властивістю обмежених опуклих множин: якщо *D* - замкнена обмежена опукла множина, що має кінцеве число кутових точок, то будь-яку точку $x \in D$ можна подати у вигляді опуклої комбінації кутових точок *D*.

Нехай $x_1, x_2, ..., x_m$ - кутові точки множини *D*, а x^* - точка, у якій цільова функція *L* досягає максимуму. Точку x^* можна уявити як опуклу комбінацію кутових точок $x_1, x_2, ..., x_m$

$$x^* = \sum_{i=1}^{m} \lambda_i x_i \ ,$$

де $\sum \lambda_i = 1, \quad \lambda_i \geq 0, \quad i = \overline{1, m}$.

Оскільки x^* - точка максимуму, то для будь-якого x $cx^* \geq cx$. У тому числі й для cx_r (x_r – кутова точка).

Функція $L(x)$ - лінійна, тому

$$L\left(\sum_{i=1}^{m} \lambda_i x_i\right) = \sum_{i=1}^{m} \lambda_i L(x_i) \, ,$$

а значить

$$cx^* = c\sum_{i=1}^{m} \lambda_i x_i = \sum \lambda_i (cx_i) \leq \sum \lambda_i (cx_r) = cx_r \sum \lambda_i = cx_r \, ,$$

де x_r - кутова точка множини *D*, що задовольняє умові

$$cx_r = \max_{1<i<m} \{cx_i\}.$$

Таким чином, $cx^* \leq cx_r$. У той самий час $cx^* \geq cx_r$, звідки випливає $cx^* = cx_r$.

Тобто існує принаймні одна кутова точка x_r, у якій цільова функція приймає максимальне значення.

Теорема 2.2 Якщо цільова функція *L* приймає максимальне значення в кількох точках множини *D*, то вона приймає це саме значення в будь-якій точці, що є їхньою опуклою комбінацією.

Доказ.

Нехай максимальне значення цільової функції $L(x)$ досягається в точках $x_1, x_2, \ldots, x_s$, тобто $L^* = cx_i$, $i = \overline{1, s}$. Розглянемо довільну опуклу комбінацію цих точок

$$x^* = \sum_{i=1}^{s} \lambda_i x_i ,$$

де $\sum \lambda_i = 1, \quad \lambda_i \geq 0, \quad i = \overline{1, s}$.

Знайдемо значення цільової функції в точці x^*

$$L(x^*) = cx^* = c\sum_{i=1}^{s} \lambda_i x_i = \sum \lambda_i (cx_i) = \sum \lambda_i L^* = L^* \sum \lambda_i = L^*.$$

Теорема 2.1 є фундаментальною, тому що вона вказує принциповий шлях розв'язання ЗЛП. Замість дослідження нескінченної множини припустимих розв'язків для знаходження серед них оптимального, необхідно досліджувати лише кінцеве число кутових точок багатогранника розв'язків.

2.3 Канонічна форма задачі лінійного програмування (КЗЛП)

Канонічною називається задача лінійного програмування, якщо всі її обмеження є рівняннями.

Переважна більшість методів розв'язання задач лінійного програмування призначена для канонічних задач. Тому початковий етап розв'язання будь-якої загальної ЗЛП завжди пов'язаний із приведенням її до еквівалентної КЗЛП.

Загальна ідея переходу від ЗЗЛП до КЗЛП досить проста. Обмеження у вигляді нерівностей перетворюють на рівняння за рахунок додавання фіктивних невід'ємних змінних, які одночасно входять до цільової функції з коефіцієнтом 0, тобто не надають впливу на її значення.

$$x'_{n+i}, \quad i = \overline{1, k}.$$

Змінні, на які не накладена умова невід'ємності, подаються у вигляді різниці двох нових невід'ємних змінних.

Додатково треба помітити, що вибір типу шуканого екстремуму (максимуму або мінімуму) носить відносний характер. Так, задача пошуку максимуму функції

$$L(x) = \sum_{j=1}^{n} c_j x_j$$

еквівалентна задачі пошуку мінімуму функції

$$-L(x) = \sum_{j=1}^{n} -c_j x_j .$$

У результаті таких перетворень одержують канонічну задачу:

Знайти найбільше або найменше значення функції

$$L = c_1x_1 + c_2x_2 + \ldots + c_nx_n + 0 \cdot x'_{n+1} + \ldots + 0 \cdot x'_{n+k} \tag{2.3}$$

на деякій множині D, де $x \in D$ задовольняє системі обмежень

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + x'_{n+1} = b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + x'_{n+2} = b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{k1}x_1 + a_{k2}x_2 + \ldots + a_{kn}x_n + x'_{n+k} = b_k, \\ a_{k+1,1}x_1 + a_{k+1,2}x_2 + \ldots + a_{k+1,n}x_n = b_{k+1}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m, \\ x_j \ge 0, \quad j = \overline{1, n+k}. \end{cases} \tag{2.4}$$

Можна показати, що отримана задача є еквівалентною вихідній задачі.

Таким чином, перехід від загальної форми задачі лінійного програмування до канонічної призводить до зростання її розмірності, що ускладнює процес розв'язання.

2.3.1 Базисні розв'язки канонічної задачі лінійного програмування

Запишемо КЗЛП у матричній формі:

знайти максимум

$$\overline{C}^T \overline{X} \tag{2.5}$$

за умови

$$\overline{A}^T \overline{X} = \overline{B}, \qquad \overline{X} \ge 0, \tag{2.6}$$

де $\overline{A}$, $\overline{X}$ і $\overline{B}$ – вектори.

$$\overline{A_1} = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}; \qquad \overline{A_2} = \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} \qquad \overline{A_n} = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}; \qquad \overline{B} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}. \tag{2.7}$$

Будемо розглядати стовпці матриці $\overline{A}_j$ як вектори простору R^m. Тоді кожному припустимому плану КЗЛП (n-мірному вектору $\overline{X}$) відповідає лінійна комбінація стовпців $\overline{A}_j$, щог дорівнює стовпцю $\overline{B} \in R^m$.

Вектори $\overline{A}_j$, $j = \overline{1, n}$ називають векторами вимог задачі. Вектор $\overline{B}$ називається вектором обмежень. При цьому компоненти певного припустимого плану $x \in D$ - x_j є коефіцієнтами розкладання вектора обмежень $\overline{B}$ задачі за векторами вимог $\overline{A}_j$.

Очевидно, що число рівнянь, які задають множину D, менше або дорівнює числу змінних задачі ($m \le n$). Якщо це не так, то або система рівнянь

$$\overline{A}\ \overline{X} = \overline{B}$$

є несумісною, або вона містить надлишкові (лінійно залежні) рівняння.

Якщо деякі m стовпців $A_1, A_2, \ldots, A_m$ матриці $\overline{A}$ є лінійно незалежними, то вони утворюють ***базис*** у просторі R^m та їх буде досить для подання вектора $\overline{B}$

у вигляді лінійної комбінації зазначених стовпців. Це означає, що інші стовпці ввійдуть у дане розкладання з нульовими коефіцієнтами.

Якщо коефіцієнти лінійної комбінації виявляться невід’ємними, то ми отримаємо **базисний припустимий план** *x*, у якого кількість елементів, що відмінні від нуля не перевищує *m*. Або іншими словами. Система обмежень КЗЛП являє собою систему m рівнянь із *n* змінними, причому $m \le n$.

У такій системі *m* змінних називаються **базисними**, а інші (*n*-*m*) змінних – **вільними**.

Базисним розв'язком системи *m* лінійних рівнянь з *n* змінними називається розв'язок, у якому всі n-m вільних змінних дорівнюють нулю.

Базисними розв'язками можуть бути різні групи з *n* змінних. У загальному випадку число таких груп не перевершує C_n^m.

Таким чином, система з *m* лінійних рівнянь із *n* змінними ($m<n$) є **невизначеною**, тому що кожному довільному набору вільних змінних відповідає один базисний розв'язок системи.

Опорним планом КЗЛП називається припустимий базисний розв'язок, компоненти якого перевищують нуль.

Базисний розв'язок, у якому хоча б одна з базисних змінних дорівнює нулю, називається **виродженим**.

2.3.2 Властивості базисних планів задачі лінійного програмування

Теорема 2.3 Кожний припустимий базисний план є кутовою точкою множини припустимих планів *D*.

Доказ.

Будемо вважати, що базисними є перші m стовпців матриці $\overline{A}$

$$A_1, A_2, \ldots, A_m \ldots$$

Тоді формулювання теореми можна записати у вигляді:

Якщо існує такий *n*-мірний вектор

$$x = \left(x_1, x_2, \ldots, x_k, \underbrace{0, \ldots, 0}_{n-k} \right),$$

$$x_j > 0, \quad j = \overline{1,k}, \quad k \le m,$$

що $x_1 A_1 + x_2 A_2 + \ldots + x_k A_k = B_0$, то *x* є кутовою точкою множини *D*.

Припустимо, що базисний план *x* не є кутовою точкою множини *D*. У цьому випадку його можна подати у вигляді опуклої комбінації двох різних припустимих планів x^1 і x^2

$$x = \lambda x^1 + (1-\lambda) x^2, \quad 0 < \lambda < 1.$$

Оскільки останні (*n*-*k*) компонент вектора *x* дорівнюють нулю, а λ і (1-λ) перевищують нуль, то ці самі (*n*-*k*) компонент у векторах x^1 і x^2 також дорівнюють нулю.

Оскільки плани x^1 і x^2 – припустимі плани,

$$x_1^1 A_1 + x_2^1 A_2 + \ldots + x_k^1 A_k = B,$$
$$x_1^2 A_1 + x_2^2 A_2 + \ldots + x_k^2 A_k = B.$$

Віднімемо з першого рівняння друге

$$(x_1^1 - x_1^2)A_1 + (x_2^1 - x_2^2)A_2 + \ldots + (x_k^1 - x_k^2)A_k = 0.$$

Отримали нульовий вектор. Оскільки A_j лінійно незалежні, нулю дорівнюють коефіцієнти

$$(x_1^1 - x_1^2) = 0,$$
$$(x_2^1 - x_2^2) = 0,$$
$$\vdots$$
$$(x_k^1 - x_k^2) = 0.$$

Звідки випливає, що $x^1 = x^2$. Це суперечить припущенню, що x^1 і x^2 є різними кутовими точками множини D.

Таким чином, план x не може бути поданий у вигляді опуклої комбінації двох інших точок множини, отже є кутовою точкою даної множини.

Справедливо й зворотне ствердження: Якщо x - кутова точка множини D, то вона є припустимим базисним планом задачі лінійного програмування.

2.4 Симплекс-метод

2.4.1 Загальна характеристика

На підставі розглянутих теорем про властивості лінійних екстремальних задач можна укласти, що пошук їхніх розв'язків збігається до послідовного перебору кутових точок множини припустимих планів.

Проте такий перебір для реальних багатомірних завдань на практиці нездійсненний або вкрай неефективний. Наприклад, число базисних планів у задачі з 10 змінними й 30 обмеженнями становить

$$C_{30}^{10} = \frac{30!}{10!*20!} = 1489411.$$

Класичним методом розв'язання задач лінійного програмування є симплекс-метод, що також називається методом послідовного поліпшення плану. Він розроблений у 1947 році американським математиком Джорджем Данцигом.

Симплекс-метод являє собою послідовний перебір кутових точок області припустимих розв'язків, при якому значення цільової функції зростає від ітерації до ітерації (від однієї кутової точки до іншої).

Критерій оптимальності в симплекс-методі реалізується шляхом визначення спеціальних оцінок Δ_j для небазисних векторів-стовпців матриці $\overline{A}$, щодо поточного базису (симплекс-різниць). Симплекс-різниці обчислюють за формулою

$$\Delta_j = L_j - c_j, \qquad (2.8)$$

де L_j – індекси векторів, що відповідають поточному базису

$$L_j = \sum_{i=1}^{m} c_j a_{ij} \ . \qquad (2.9)$$

Сформулюємо критерій оптимальності припустимого базисного плану:

план є оптимальним, якщо для всіх $j = \overline{1,n}$ $\Delta_j \geq 0$, і неоптимальним у протилежному випадку, тобто якщо існує таке $j = \overline{1,n}$, що $\Delta_j < 0$.

Якщо симплекс-різниці показують неоптимальність плану, здійснюється перехід до наступного базису. При цьому один стовпець виводиться з базису, а інший вводиться. Для забезпечення поліпшення значення цільової функції в базис повинен бути введений вектор-стовпець, що має від'ємну оцінку. Якщо таких стовпців декілька, то для уведення рекомендується вибирати стовпець, що має максимальний за модулем добуток оцінки Δ_j на відношення $\Theta_r = \min\limits_i \left\{ \frac{b_i}{a_{ij}} \right\}$. Одночасно на цьому кроці потрібно ухвалити рішення щодо того, який стовпець треба вивести з базису. Зробити це потрібно таким чином, щоб знову формований базис опинився припустимим.

Можна довести, що припустимість базису, до якого здійснюється перехід, забезпечується наступним правилом виводу стовпця з поточного базису:

для стовпця, що претендує на введення до базису, і вектора обмежень $\overline{B}$ розглядаються відносини

$$\Theta_i = \frac{b_i}{a_{ij}} . \qquad (2.10)$$

і визначається такий рядок r, що

$$\Theta_r = \min_i \{\Theta_i\}. \qquad (2.11)$$

Отриманий індекс r визначає номер рядка, який відповідає вектору, виведеному з базису.

Таким чином, якщо базис на q-ї ітерації включав стовпці з номерами

$$\{j_1, j_2,.., j_{r-1}, j_r, j_{r+1},.., j_m\},$$

то базис на ітерації $q + 1$ складатиметься зі стовпців з номерами:

$$\{j_1, j_2,.., j_{r-1}, j_{r+1},.., j_m\}.$$

Окремо слід домовитися про випадок, коли стовпець, що претендує на введення до базису, не містить додатних компонентів (a_{ij}<0). Це означає, що цільова функція в задачі не обмежена на множині припустимих значень, тобто може досягати як завгодно великого значення, отже оптимальний план відсутній.

Після переходу до нового базису можна заново сформувати матрицю $\overline{A}$ й дослідити новий план на оптимальність. З погляду техніки обчислень представляється раціональним безпосередньо переходити від матриці q-ї ітерації до матриці (q+1)-ї ітерації. Справа в тому, що в цих матриць стовпці, які відповідають базисним векторам, складаються з нулів, за винятком одного елемента, що дорівнює одиниці. Позиція цього ненульового (одиничного) елемента визначається порядковим номером базисного стовпця. Тому для одержання матриці (q+1)-ї ітерації досить за допомогою лінійних операцій над рядками матриці q-ї (попередньої) ітерації привести її стовпець, що відповідає вектору, який вводиться до базису, до «базисного» вигляду.

Для цього застосовується перетворення Жордана-Гауса (метод повного виключення). У цьому випадку воно полягає в тому, що треба дістати одиницю на місці елемента a_{rj} (він зазвичай називається ведучим) і нулі на місці інших елементів стовпця a_{ij}. Перше досягається за допомогою ділення r-го рядка на ведучий елемент, друге - шляхом додавання знову отриманого r-го рядка, помноженого на відповідний коефіцієнт, до інших рядків матриці q-ї (попередньої) ітерації.

Формально результат виконання даного перетворення над елементами матриці можна виразити у наступному виді:

$$a_{rj}^{q+1} = \frac{a_{rj}^{q}}{a_{rk}^{q}}, \qquad b_{r}^{q+1} = \frac{b_{r}^{q}}{a_{rk}^{q}}, \tag{2.12}$$

де $j = \overline{1,n}$;

k – номер стовпця, що вводиться до базису;

$$a_{ij}^{q+1} = a_{ij}^{q} - a_{ik}^{q}\,\frac{a_{rj}^{q}}{a_{rk}^{q}}, \tag{2.13}$$

де $i = \overline{1,m}$, $i \neq r$;

$$b_{i}^{q+1} = b_{i}^{q} - a_{ik}^{q}\,\frac{b_{r}^{q}}{a_{rk}^{q}}, \tag{2.14}$$

де $i = \overline{1,m}$, $i \neq r$.

Слід особливо відзначити зміст елементів вектора $\overline{B}$. Його нульовий компонент b_0 містить значення цільової функції, що досягається нею на поточному плані, а інші елементи — ненульові компоненти цього плану.

Приведемо схему алгоритму симплекс-методу для розв'язання задачі максимізації.

1. Визначення припустимого базисного плану.

2. Перевірка оптимальності поточного базисного плану: здійснюється перегляд рядка оцінок Δ_j. Можливі два варіанти:

- $\Delta_j \geq 0$ - план, що відповідає поточному базису задачі, є оптимальним. Обчислювальний процес закінчений. Виписується оптимальний план задачі x^* і значення цільової функції $L(x^*)$.

- у рядку оцінок Δ_j існує щонайменше один елемент $\Delta_j<0$, тобто оцінка якого є від’ємною. Отже, поточний план є неоптимальним. Вибирається стовпець із номером k, для якого добуток $\Delta_j\ \Theta$ найбільший за абсолютною величиною. Він називається ведучим і має бути введений до чергового базису.

3. Визначення стовпця, що виводитиметься з базису. Досліджується ведучий стовпець, можливі два варіанти:

- для всіх $i = \overline{1,m}$ $a_{ik}^{q}<0$. Робиться висновок про необмеженість цільової функції та обчислювальний процес завершується;

- існує принаймні один рядок з номером $i=\overline{1,m}$, для якого $a_{ik}^{q}>0$. Відповідно до правила (2.11) визначається номер r стовпця, що виводитиметься з базису.

4. Перерахування елементів матриці $\overline{A}$ й стовпця $\overline{B}$ щодо нового базису відповідно до формул (2.12)-(2.14). Перехід до пункту 2 алгоритму.

2.4.2 Таблична реалізація симплекс-методу

З погляду забезпечення раціональності й наочності обчислювального процесу виконання алгоритму симплекс-методу зручно оформляти у вигляді послідовності таблиць. У різних джерелах наводяться різні модифікації симплекс-таблиць, що відрізняються одна від одної розташуванням окремих елементів. Проте всі вони базуються на тих самих принципах. Зупинимося на структурі таблиці, що наведена у таблиці 2.1.

Таблиця 2.1 - Структура симплексної таблиці

Базис	$C_{баз}$	C_j	-6	-5	0	0	0
		B	A_1	A_2	A_j	**…**	A_n
A_j							
	L_j Δ_j						

Стовпець «Базис» містить найменування базисних векторів (у тій послідовності, у якій вони входять до базису), стовпець $C_{баз}$ містить коефіцієнти цільової функції при базисних змінних, стовпець B - компоненти вектора обмежень щодо поточного базису, A_1-A_n - компоненти матриці задачі щодо поточного базису. У рядку L_j записують індекси, що визначені за формулою (2.9).

У рядку Δ_j знаходяться поточні оцінки стовпців. Рядок C_j містить коефіцієнти при компонентах поточного плану в цільовій функції.

Треба відзначити, що таблична модифікація симплекс-методу має важливе практичне значення не стільки як зручна форма організації ручного рахунку, скільки як основа для реалізації даного алгоритму в рамках програмного забезпечення ЕОМ.

Розглянемо приклад розв'язання ЗЛП симплекс-методом. Нехай дана канонічна задача ЛП:

$$L(x)= 50x_1 - 10x_2 + 6x_3 + 40x_4 - 30x_5 \to \max,$$
$$x_1 + x_4 + 3x_5 = 12,$$
$$2x_1 + x_2 + 3x_4 = 14,$$
$$-2x_1 + 3x_3 - 4x_4 = 17,$$
$$x_1>0,\ x_2>0,\ x_3>0,\ x_4>0,\ x_5>0.$$

Як видно, стовпці матриці з номерами {5, 2, 3} є лінійно незалежними, і можна одержати розкладання за даними стовпцями вектора обмежень із додатними коефіцієнтами. Останнє означає, що стовпці {5, 2, 3} утворюють припустимий базис, з якого можна почати розв'язання задачі. Початковий опорний план має вигляд x^1 = {0, 14, 17/3, 0, 4}. Заповнюємо симплекс-таблицю 2.2, що відповідає першій ітерації (q=1).

Таблиця 2.2 - Перша ітерація

Базис	$C_{баз}$	C_j	50	-10	6	40	-30
		B	A_1	A_2	A_3	A_4	A_5
A_2	-10	14	2	1	0	3	0
A_3	6	17/3	-2/3	0	1	-4/3	0
A_5	-30	4	1/3	0	0	1/3	1
	L_j	-226	-34	-10	6	-48	-30
	Δ_j		-84	0	0	-88	0

Оскільки рядок оцінок у першому й четвертому стовпцях містить від'ємні елементи Δ_1 = -84, Δ_4= -88, план x^1 = {0, 14, 17/3, 0, 4} не є оптимальним, і значення цільової функції $L(x^1)$ = -226 можна покращити. Перейдемо до нового опорного плану.

Визначимо вектор, що будемо вводити до базису (A_1 або A_4).

Відповідно до алгоритму симплекс-методу визначимо відношення Θ. Для вектора A_1: $\Theta_r = \min_i \left\{ \frac{14}{2}; \frac{4}{1/3} \right\} = \min\{7;12\} = 7$, r=2; для вектора A_4: $\Theta_r = \min_i \left\{ \frac{14}{3}; \frac{4}{1/3} \right\} = \min\{4{,}66;12\} = 4{,}66$, r=2. Добутки для вектора A_1: -34*7= -238; для вектора A_4: -48*4,66= -223,7. Вважаємо номер стовпця, що вводиться в черговий базис, k = 1 (тому що |-238|>|-223,7|). З базису потрібно вивести стовпець із номером 2. Отримуємо черговий припустимий базис задачі {1, 3, 5}. Елемент, що знаходиться на перетинанні виділених стовпця й рядка таблиці є ведучим, він дорівнює 2. Застосувавши формули (2.12)-(2.14), переходимо до симплекс-таблиці, що відповідає другій ітерації, і вважаємо індекс поточної ітерації q = 2.

Таблиця 2.3 - Друга ітерація

Базис	$C_{баз}$	C_j	50	-10	6	40	-30
		B	A_1	A_2	A_3	A_4	A_5
A_1	50	7	1	1/2	0	3/2	0
A_3	6	10,3	0	0,2	1	-0,3	0
A_5	-30	1,7	0	-0,2	0	-0,2	1
	L_j	360,8	50	32,2	6	79,2	-30
	Δ_j		0	42,2	0	39,2	0

Одержуємо новий план x^2 ={7; 0; 10,3; 0; 1,7}. Як видно з таблиці, рядок оцінок містить тільки невід'ємні значення, тому досягнутий базис {1, 3, 5} є оптимальним. Отже, вектор x^*={7; 0; 10,3; 0; 1,7} є оптимальним планом (точкою максимуму) задачі, максимальне значення цільової функції дорівнює $L^* = L(x^*)$ = 360,8.

2.4.3 Збіжність симплекс-методу. Виродженість у задачах ЛП

Найважливішою властивістю будь-якого обчислювального алгоритму є збіжність, тобто можливість одержання в ході його застосування шуканих результатів (із заданою точністю) за кінцеве число кроків (ітерацій).

Легко помітити, що проблеми зі збіжністю симплекс-методу потенційно можуть виникнути на етапі вибору значення r у випадку, коли однакові мінімальні значення відношення $\Theta_r = \min_i \left\{ \frac{b_i}{a_{ij}} \right\}$ будуть досягнуті для кількох рядків таблиці одночасно. Тоді на наступній ітерації стовпець b^{q+1}-й буде містити нульові елементи. Нагадаємо, що припустимий базисний план канонічної задачі ЛП, що відповідає поточному базису, називається виродженим, якщо деякі його базисні компоненти дорівнюють нулю, тобто вектор b^q містить нульові елементи.

Задача ЛП, що має вироджені плани, називається виродженою. При виході на вироджений план ми фактично одержуємо розкладання вектора $\overline{B}$ на меншу за m кількість векторів $\overline{A}_j$ і, отже, втрачаємо можливості коректно визначити, введення якого стовпця до базису призведе до зростання значення цільової функції. Подібні ситуації, в остаточному підсумку, можуть привести до зациклення обчислювального процесу, тобто нескінченному перебору тих самих базисів.

З погляду першої геометричної інтерпретації ЗЛП ситуація виродженості означає, що через певну кутову точку багатогранної множини припустимих планів задачі, що відповідає поточному базисному плану, проходить більше за m гіперплощин обмежень задачі. Іншими словами, одне або кілька обмежень у даній точці є надлишковими. Останнє дає повід для міркувань про економічний бік проблеми виродженості як проблеми наявності надлишкових обмежень і в деяких випадках визначає шляхи її розв'язання.

Ідея методу розв'язання вироджених задач ЛП, що дістала назви **методу збурень**, полягає в тому, що при виході на вироджений план здійснюють незначний зсув вектора $\overline{B}$, і виродженість усувається.

Необхідно сказати, що розглянута проблема зациклення для більшості практично значущих задач є досить рідкою й малоймовірною. Більше того, вона дуже часто вирішується за рахунок помилок округлень при виконанні розрахунків на ЕОМ.

Знаходження припустимого базисного плану. У розглянутому вище прикладі вихідний базисний план, необхідний для початку обчислень за симплекс-методом, був підібраний за рахунок особливостей матриці умов. Дійсно, дана матриця вже містила необхідну кількість «майже базисних» стовпців. Очевидно, що для переважної більшості задач ЛП неможливо подібним чином відразу й у явному вигляді вказати вихідний припустимий базисний план. Існують різні прийоми розв'язання даної задачі. Ми зупинимося на одному з них, що отримав назву **методу мінімізації нев'язань**. Його сильною стороною, безумовно, є універсальність. Хоча, у деяких окремих випадках, він може виявитися занадто громіздким.

Ідея методу мінімізації нев'язань полягає в побудові відповідної допоміжної задачі, для якої можна в явному вигляді вказати вихідний базисний план, і розв'язанні її за допомогою процедури симплекс-методу.

Можна вважати, що вектор обмежень у вихідній задачі невід'ємний, тобто $b_i > 0, \quad i = \overline{1, m}$. Тоді для КЗЛП максимізації функції

$$L = c_1x_1 + c_2x_2 + \ldots + c_nx_n \to \max \tag{2.15}$$

на множині, що зумовлена обмеженнями

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m, \\ x_j \geq 0, \quad j = \overline{1, n} \end{aligned} \tag{2.16}$$

будується допоміжна задача

$$L = 0x_1 + 0x_2 + \ldots + 0x_n - 1\,x_{n+1} - \ldots - 0 \cdot x_{n+m}. \tag{2.17}$$

На новій множині $D' = \{x \in R^{n+m}\}$, що зумовлена системою обмежень

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + 1 * x_{n+1} + \ldots + 0 * x_{n+m} &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + 0 * x_{n+1} + 1 * x_{n+2} + \ldots + 0 * x_{n+m} &= b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n + 0 * x_{n+1} + 0 * x_{n+2} + \ldots + 1 * x_{n+m} &= b_m, \\ x_j \geq 0, \quad j = \overline{1, n+m}. \end{aligned} \tag{2.18}$$

Як видно, допоміжна задача (2.17)-(2.18) відрізняється від основної задачі (2.15)-(2.16) матрицею, що отримана в результаті додавання до вихідної матриці $\overline{A}$ одиничної матриці. Їй відповідають додаткові (фіктивні) змінні $x_{n+1},\ldots,x_{n+m}$. Ці змінні (їх називають нев'язаннями) у цільовій функції мають коефіцієнти (-1), а змінні $x_1,\ldots,x_n$, які є загальними для обох задач, входять до цільової функції з нульовими коефіцієнтами. Істотною особливістю допоміжної задачі є наявність в неї очевидного вихідного базису, утвореного доданими стовпцями, і відповідного плану з базисними компонентами $x_{n+i} = b_i \geq 0$, $i = \overline{1, m}$. У силу структури цільової функції в процесі пошуку її максимуму процедурою симплекс-методу фіктивні змінні (нев'язання) x_{n+i} повинні мінімізуватися, бажано до нуля, чого можна досягнути за рахунок послідовного переведення їх у небазисні компоненти плану. Якщо в оптимальному плані x^*, отриманому в результаті розв'язання допоміжної задачі, останні m компонент (тобто нев'язання) дорівнюють нулю, то його початкові n компонент задовольняють системі обмежень, що визначають область D, і x^* простим відкиданням нев'язань перетворюють на припустимий план основної задачі. При цьому всі рядки симплекс-таблиці, отриманої на останній ітерації допоміжної задачі, переносять до першої таблиці основної задачі.

У випадку, коли оптимальний план допоміжної задачі все-таки містить фіктивні компоненти, що не дорівнюють нулю, припустимі плани у вихідної задачі відсутні, тобто $D = \varnothing$.

Контрольні запитання

1. Сформулюйте задачу лінійного програмування.
2. Дайте визначення для наступних понять: план, припустимий план, оптимальний план, розв'язок задачі.
3. Чим відрізняється загальна задача лінійного програмування від канонічної?
4. Чи завжди загальну задачу лінійного програмування можна призвести до канонічного вигляду?
5. Яка точка опуклої множини називається кутовою?
6. У чому полягає перша геометрична інтерпретація задачі лінійного програмування?
7. Який план ЗЛП називається базисним?
8. Як зв'язані базисні плани та кутові точки області визначення задачі лінійного програмування?
9. Який план задачі лінійного програмування називається виродженим?
10. Сформулюйте критерій оптимальності припустимого базисного плану, застосовуваний у симплекс-методі.
11. Сформулюйте основні етапи стандартної ітерації симплекс-методу.
12. Для чого застосовується перетворення Жордана-Гаусса?
13. Який елемент симплекс-таблиці називається ведучим?
14. За яких умов робиться висновок про необмеженість цільової функції в розв'язуваній задачі?
15. Чи можна заздалегідь точно визначити кількість ітерацій, що буде потрібна для розв'язання задачі симплекс-методом? Чи можна знайти верхню границю для даної величини?
16. Яка задача називається виродженою? За якими ознаками можна впізнати, що поточний план є виродженим?
17. У чому полягає основна ідея методу збурень?
18. Для чого призначений метод мінімізації нев'язань?

ТЕМА 3
ТРАНСПОРТНА ЗАДАЧА

3.1 Транспортна задача в матричній постановці та її властивості

Оптимізаційна модель задачі. Нехай є m пунктів відправлення (постачальників) A_1, A_2, ... A_m, у яких знаходиться однорідна продукція в кількостях a_1, a_2, ..., a_m відповідно. Нехай є n пунктів призначення (споживачів) B_1, B_2, ..., B_n, що подали заявки відповідно на b_1, b_2, ..., b_n одиниць вантажу. Сума всіх заявок дорівнює сумі всіх запасів.

Відомі вартості c_{ij} перевезення одиниці продукції з i-го пункту відправлення в j-й пункт призначення. Вважається, що вартість перевезення кількох одиниць вантажу пропорційна їхньому числу.

Потрібно скласти такий план перевезень (звідки, куди й скільки одиниць везти), щоб всі заявки були виконані, а загальна вартість всіх перевезень була мінімальною.

Поставимо цю задачу як задачу лінійного програмування. Позначимо x_{ij} - кількість одиниць вантажу, що відправляється з i-го ПВ А в j-й ПП В.

Сукупність чисел x_{ij} ми будемо називати «планом перевезень», а самі величини x_{ij} - «перевезеннями». Ці невід’ємні змінні повинні задовольняти наступним умовам:

1. Сумарна кількість вантажу, що направляється з кожного ПВ в усі ПП, повинна дорівнювати запасу вантажу в даному пункті.

2. Сумарна кількість вантажу, що доставляється до кожного ПП з усіх ПВ, повинна дорівнювати заявці, поданої даним пунктом.

3. Сумарна вартість всіх перевезень, тобто сума величин x_{ij}, помножених на відповідні вартості c_{ij}, повинна бути мінімальною.

Таким чином, дана задача збігається до визначення такого плану перевезень певного продукту з пунктів його виробництва до пунктів споживання $x = (x_{11}, \ldots, x_{1n}, x_{21}, \ldots, x_{2n}, \ldots, x_{i1}, \ldots, x_{in}, \ldots, x_{m1}, \ldots, x_{mn})$, який мінімізує цільову функцію

$$f(x) = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \tag{3.1}$$

на множині припустимих планів

$$\sum_{j=1}^{n} x_{ij} = a_i, \ \sum_{i=1}^{m} x_{ij} = b_j, \ x_{ij} \geq 0, \ i = \overline{1,m}, \quad j = \overline{1,n} \tag{3.2}$$

за умови дотримання балансу

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j \ . \tag{3.3}$$

Якщо умова (3.3) виконується, то задача називається **збалансованою** або **закритою**, а інакше задача називається **незбалансованою** або **відкритою**.

Транспортна задача є представником класу задач лінійного програмування і тому має всі властивості лінійних оптимізаційних задач, але одночасно вона має ще низку додаткових корисних властивостей, які дозволили розробити спеціальні методи її розв'язання.

Кількість лінійно незалежних серед рівнянь в умовах-обмеженнях ТЗ дорівнює

$$m+n-1,$$

отже кількість базисних змінних так само дорівнює

$$m+n-1.$$

Загальне число змінних x_{ij} у транспортній задачі дорівнює $m * n$, а число вільних змінних

$$k = mn - (m + n - 1) = (m - 1) (n - 1).$$

Відомо, що в задачі лінійного програмування оптимальний розв'язок досягається в одній з вершин ОДР, в опорній точці, де принаймні k змінних дорівнюють нулю. Виходить, у нашому випадку для оптимального плану принаймні $(m-1)(n-1)$ перевезень повинні дорівнювати нулю (з відповідних ПВ у відповідні ПП нічого не перевозиться).

Будемо називати будь-який план перевезень **припустимим**, якщо він задовольняє умовам-обмеженням (всі заявки задоволені, всі запаси вичерпані).

Припустимий план будемо називати **опорним**, якщо в ньому відмінні від нуля не більше за $m + n - 1$ базисних перевезень, а інші $(m-1)(n-1)$ перевезень дорівнюють нулю.

План будемо називати **оптимальним**, якщо він, серед всіх припустимих планів, призводить до мінімальної сумарної вартості перевезень (L=min).

У силу особливої структури ТЗ при її розв'язанні не потрібно розв'язувати систему рівнянь. Всі операції із знаходження оптимального плану збігаються до маніпуляцій безпосередньо з таблицею, де в певному порядку записані умови транспортної задачі: перелік ПВ та ПП, заявки та запаси, а також вартості перевезень c_{ij}. У міру заповнення цієї таблиці в її клітках проставляються самі перевезення x_{ij}. Транспортна таблиця складається з m рядків та n стовпців. Рядки транспортної таблиці відповідають пунктам виробництва (в останній клітці кожного рядка зазначений обсяг запасу продукту a_i), а стовпці - пунктам споживання (остання клітка кожного стовпця містить значення заявки b_j). У правому верхньому куті кожної клітки ставлять вартість c_{ij} перевезення одиниці продукту з A_i у B_j, а центр клітки залишають вільним, щоб поміщати у нього саме перевезення x_{ij}. Клітки, які містять нульові перевезення (x_{ij}=0), називають вільними, а ненульові - зайнятими (x_{ij}>0).

3.2 Методи побудови опорного плану

За аналогією з іншими задачами лінійного програмування розв'язання транспортної задачі починається з побудови припустимого базисного плану. Найбільш простий спосіб його знаходження ґрунтується на так званому **методі північно-західного кута**. Суть методу полягає в послідовному розподілі всіх запасів, наявних у першому, другому й т.д. пунктах виробництва, до першого, другого й т.д. пунктів споживання. Кожний крок розподілу зводиться до спроби повного вичерпання запасів у черговому пункті виробництва або до спроби повного задоволення потреб у черговому пункті споживання. На кожному кроці q величини поточних нерозподілених запасів позначаються a_i^q, а поточних незадоволених потреб - b_j^q. Побудова припустимого початкового плану, відповідно до методу північно-західного кута, починається з лівого верхнього кута транспортної таблиці. Для чергової клітки, розташованої в рядку і і стовпці j, розглядаються значення нерозподіленого запасу в і-ому пункті виробництва та незадоволеної заявки в j-ому пункті споживання, з них вибирається мінімальне та призначається як обсяг перевезення між даними пунктами: $x_{ij} = \min\{a_i^q, b_j^q\}$. У результаті цього значення нерозподіленого запасу та незадоволеної потреби у відповідних пунктах зменшуються:

$$a_i^{(q+1)} = a_i^q - x_{ij}; \qquad b_j^{(q+1)} = b_j^q - x_{ij}.$$

Очевидно, що на кожному кроці виконується хоча б одна з рівностей: $a_i^{(q+1)} = 0$ або $b_j^{(q+1)} = 0$. Якщо справедливо $a_i^{(q+1)} = 0$, то це означає, що весь запас і-го пункту виробництва вичерпаний і необхідно перейти до розподілу запасу в пункті виробництва і+1, тобто переміститися до наступної клітки вниз по стовпцю. Якщо ж $b_j^{(q+1)} = 0$, то значить повністю задоволена заявка для j-го пункту, після чого виконується перехід на клітку, розташовану праворуч по рядку. Знову обрана клітка стає поточною, і для неї повторюються всі перелічені операції.

Ґрунтуючись на умові балансу запасів і заявок (3.3), неважко довести, що за кінцеве число кроків буде отриманий припустимий план. У силу тієї ж умови число кроків алгоритму не може бути більшим за $m+n-1$, тому завжди залишаться вільними (нульовими) $mn-(m+n-1)$ кліток. Отже, отриманий план є базисним. Не виключено, що на деякому проміжному кроці поточний нерозподілений запас виявиться рівним поточній незадоволеній заявці ($a_i^q = b_j^q$). У цьому випадку перехід до наступної клітки відбувається в діагональному напрямку (одночасно змінюються поточні пункти виробництва й споживання), а це означає «втрату» одного ненульового компонента в плані або, інакше кажучи, виродженість побудованого плану.

Розглянемо приклад. З 3-х пунктів виробництва необхідно вивезти однорідний продукт в 5 пунктів споживання. Транспортні витрати, обсяг виробництва й споживання наведені в таблиці 3.1.

Таблиця 3.1 - Вихідні дані

	B_1	B_2	B_3	B_4	B_5	**Запаси**
A_1	7	5	2	8	7	125
A_2	8	9	4	6	9	60
A_3	5	1	9	2	3	115
Заявки	30	50	100	40	80	300

Зауважимо, що запаси дорівнюють заявкам. Отже, задача є збалансованою.

Визначимо кількість базисних змінних

$$m + n - 1 = 3 + 5 - 1 = 7.$$

Кількість вільних змінних

$$(n\text{-}1)*(m\text{-}1) = (5\text{-}1)*(3\text{-}1) = 8.$$

Іншими методами визначення початкового опорного плану є метод мінімальної вартості, метод подвійної переваги або метод апроксимації Фогеля.

У таблиці 3.2 показаний процес пошуку припустимого плану за **методом північно-західного кута**, включаючи послідовну зміну обсягу нерозподілених запасів і незадоволених потреб. Стрілки відображають траєкторію переходу по клітках транспортної таблиці, а цифри, що знаходяться за її межами, - поточні нерозподілені залишки після призначення обсягу для чергової клітки.

Таблиця 3.2 - Визначення опорного плану за методом північно-західного кута

	B_1	**B_2**	**B_3**	**B_4**	**B_5**	**Запаси**	
A_1	30 ⇒	50 ⇒	45 ⇓			125	95 45 0
A_2			55 ⇒	5 ⇓		60	5 0
A_3				35 ⇒	80	115	80 0
Заявки	30	50	100	40	80	300	
	0	0	55				
			0	35			
				0	0		

Знайдений опорний план x = (30, 50, 45, 0, 0, ...). Значення цільової функції при цьому плані перевезень

$$L = 30*7+50*5+45*2+55*4+5*6+35*2+80*3 = 1110.$$

Особливістю припустимого плану, побудованого за методом північно-західного кута, є те, що цільова функція на ньому приймає значення, як правило, віддалене від оптимального. Це відбувається тому, що при його побудові ніяк не враховуються значення c_{ij}. У зв'язку з цим на практиці для одержання вихідного плану використовується інший спосіб - метод мінімального елемента, у якому при розподілі обсягів перевезень у першу чергу займаються клітки з найменшими цінами.

Метод мінімальної вартості полягає в тому, що в таблиці вартостей вибирають найменшу, і в цій клітці записують найменше з чисел (a_i, b_j), таблиця 3.3.

Таблиця 3.3 - Визначення опорного плану за методом мінімальної вартості

	B_1	**B_2**	**B_3**	**B_4**	**B_5**	**Запаси**
A_1	7 **25**	5	2 **100**	8	7	125
A_2	8 **5**	9	4	6	9 **55**	60
A_3	5	1 **50**	9	2 **40**	3 **25**	115
Потреби	30	50	100	40	80	300

Значення цільової функції при цьому плані перевезень

$$L = 25*7+100*2+5*8+55*9+50*1+40*2+25*3 = 1115.$$

Метод подвійної переваги показаний у таблиці 3.4. Він полягає в тому, що в кожному стовпці відзначають знаком V клітку з найменшою вартістю, потім те ж роблять у кожному рядку. У клітки з W заносять самі більші обсяги перевезень. Потім розподіляють перевезення по клітках, відзначеним V. У частині таблиці, що залишилася, перевезення розподіляють за методом найменшої вартості.

Таблиця 3.4 - Визначення опорного плану за методом подвійної переваги

	B_1	**B_2**	**B_3**	**B_4**	**B_5**	**Запаси**
A_1	7 **25**	5	2 **100** W	8	7	125
A_2	8 **5**	9	4 V	6	9 **55**	60
A_3	5 V	1 **50** W	9	2 **40** V	3 **25** V	115
Заявки	30	50	100	40	80	300

Отриманий опорний план збігається з планом, що отриманий за методом мінімальної вартості, L = 1115.

Приймати як опорний треба той план, для якого транспортні витрати виявилися найменшими. Отже, як опорний треба прийняти план, отриманий за методом північно-західного кута.

Цей план є **припустимим**, оскільки суми за рядками дорівнюють запасам, а суми за стовпцями дорівнюють заявкам.

Отриманий план є **опорним**, тому що число ненульових перевезень дорівнює $m + n - 1 = 3 + 5 - 1 = 7$, а число нульових перевезень дорівнює $(n\text{-}1)*$ $*(m\text{-}1)$ =(5-1)*(3-1) = 8.

План можна покращити (табл. 3.2), якщо зменшити перевезення в дорогій клітці, наприклад (1.1), і збільшити в дешевій (3.1).

Щоб при цьому план залишався опорним, необхідно одну з вільних кліток зробити базисною, а одну з базисних - вільною.

3.3 Метод потенціалів

Одним з методів розв'язання транспортної задачі є метод, що одержав назву **методу потенціалів**. Уперше він був запропонований в 1949 році Л. В. Канторовичем і М. К. Гавуріним. Пізніше на базі загальних ідей лінійного програмування аналогічний метод був запропонований Дж. Данцигом.

Так само як транспортна задача є окремим випадком задачі ЛП, так і метод потенціалів, загалом кажучи, може трактуватися як різновид симплексних процедур. Він являє собою ітеративний процес, на кожному кроці якого розглядається певний поточний базисний план, перевіряється його оптимальність, і при необхідності здійснюється перехід до кращого базисному плану.

Алгоритм методу потенціалів починається з вибору певного припустимого базисного плану. Якщо початковий опорний план має ***m*+*n*-1** додатних перевезень, то він називається **невиродженим**. Якщо ж опорний план має менше за ***m*+*n*-1** додатних перевезень, то він називається **виродженим**.

У цьому початковому опорному плані кожному пункту ставлять у відповідність певне число, що називається його **попереднім потенціалом**. Якщо даний план **невироджений** (число ненульових базисних кліток дорівнює m+n-1) то за ним можна так визначити потенціали u_i та v_j, щоб для кожної базисної клітки (тобто для тієї, в якій $x_{ij}>0$) виконувалася умова

$$v_j - u_i = c_{ij}, \quad \text{якщо } x_{ij} > 0. \qquad (3.4)$$

Теорема 3.1 Якщо план ТЗ є оптимальним, то йому відповідає система з $m+n$ чисел, що задовольняють умовам:

$$u_i + v_j = c_{ij} \quad \text{для} \quad x_{ij} > 0,$$

$$u_i + v_j \le c_{ij} \quad \text{для} \quad x_{ij} = 0, \qquad i = \overline{1,m}; \quad j = \overline{1,n}.$$

де u_i і v_j - потенціали постачальників і споживачів відповідно.

Отже, якщо хоча б для однієї вільної клітки

$$\Delta_{ij} = u_i + v_j - c_{ij} > 0, \qquad (3.5)$$

то план не оптимальний і вимагає покращення.

Оскільки система (3.4) містить $m+n$-1 рівнянь та $m+n$ невідомих, то один з потенціалів можна задати довільно (наприклад, дорівняти v_j або u_i, до нуля). Після цього інші невідомі u_i і v_j визначаються однозначно.

Розглянемо процес визначення потенціалів поточного плану транспортної задачі на прикладі. У таблиці 3.3 переписані умови задачі з таблиці 3.1 та її припустимий базисний план, побудований за методом північно-західного кута з таблиці 3.2.

Потенціал першого пункту виробництва приймаємо рівним нулю (u_1=0). Тепер, знаючи його, можна визначити потенціали для всіх пунктів споживання, зв'язаних з першим пунктом виробництва ненульовими перевезеннями. У цьому випадку їх три (це перший, другий і третій пункти), отримуємо:

$$v_1 = c_{11} - u_1 = 7 - 0 = 7; \; v_2 = c_{12} - u_1 = 5 - 0 = 5; \; v_3 = c_{13} - u_1 = 2 - 0 = 2.$$

Маючи v_3, та з огляду на те, що в другому рядку таблиці існують ненульові компоненти x_{23} і x_{24}, можна визначити $u_2 = c_{23} - v_3 = 4 - 2 = 2$, $v_4 = c_{24} - u_2 = 6 - 2 = 4$, після чого з'являється можливість розрахувати $u_3 = c_{34} - v_4 = 2 - 4 = -2$ і, нарешті, $v_5 = c_{35} - u_3 = 3 - (-2) = 5$. У результаті отримуємо повну систему потенціалів, показану в таблиці 3.5.

Таблиця 3.5 - Визначення потенціалів для початкового опорного плану

		v_1=7	v_2=5	v_3=2	v_4=4	v_5=5	
		B₁	**B₂**	**B₃**	**B₄**	**B₅**	**Запаси**
u_1=0	**A₁**	7 30	5 50	2 45	8	7	125
u_2=2	**A₂**	8	9	4 55	6 5	9	60
u_3=-2	**A₃**	5	1	9	2 35	3 80	115
	Заявки	30	50	100	40	80	300

Для вільних кліток транспортної таблиці обчислюють величини $\Delta_{ij}=v_j+u_i- -c_{ij}$. У таблиці 3.6 вони виписані для всіх небазисних кліток під цінами.

Таблиця 3.6 - Перевірка оптимальності поточного плану (обчислення оцінок Δ_{ij})

		$v_1=7$	$v_2=5$	$v_3=2$	$v_4=4$	$v_5=5$	
		B₁	**B₂**	**B₃**	**B₄**	**B₅**	**Запаси**
$u_1=0$	**A₁**	7 30	5 50	2 45	8 0	7 -2	125
$u_2=2$	**A₂**	8 1	9 -2	4 55	6 5	9 -2	60
$u_3=-2$	**A₃**	5 0	1 2	9 -9	2 35	3 80	115
	Заявки	30	50	100	40	80	300

Відповідно до теореми 3.1, якщо всі $\Delta_{ij}\leq 0$, то план оптимальний, у протилежному випадку, якщо існує хоча б одна клітка, для якої $\Delta_{ij} > 0$, то його можна покращити. Процес «покращення» плану полягає у визначенні клітки, що вводиться, та клітки, що виводиться. У цьому простежується змістовна аналогія методу з відповідними пунктами симплекс-процедур.

Кандидатом на введення може бути будь-яка клітка, у якій $\Delta_{ij} > 0$, оскільки після введення її до базису буде забезпечена рівність $v_j + u_i = c_{ij}$. Для визначеності рекомендується брати ту клітку, у якій оцінка Δ_{ij} максимальна. У розглянутому нами прикладі це буде клітка (3, 2).

Виведена клітка визначається за допомогою так званого ланцюжка перетворення плану, що описує характер перерозподілу вантажних потоків. У відповідності з властивостями транспортної задачі для невиродженого базисного плану в поточній таблиці можна утворити замкнутий ланцюжок, що складається з вертикальних і горизонтальних ланок, однією з вершин якого є обрана вільна клітка, а інші - зайняті клітки. У таблиці 3.7 показаний ланцюжок перетворення поточного плану щодо клітки, яка вводиться до нього, (3, 2).

Таблиця 3.7 - Перетворення поточного плану

		$v_1=7$	$v_2=5$	$v_3=2$	$v_4=4$	$v_5=5$	
		B₁	**B₂**	**B₃**	**B₄**	**B₅**	**Запаси**
$u_1=0$	**A₁**	7 30	5 50 -Θ	2 45 +Θ	8 4	7 5	125
$u_2=2$	**A₂**	8 5	9 3	4 55 -Θ	6 5 +Θ	9 3	60
$u_3=-2$	**A₃**	5 9	1 +Θ 7	9 4	2 35 -Θ	3 80	115
	Заявки	30	50	100	40	80	300

Логіка алгоритму побудови ланцюжка досить проста: «вийшовши» з клітки (3, 2) у горизонтальному напрямку, ми повинні «зупинитися» у тій зайнятій клітці плану, з якої зможемо рухатися далі за вертикаллю. У даному прикладі цій вимозі задовольняють як клітка (3, 4), так і клітка (3, 5). Проте ланцюжок від (3, 5) не може бути продовжений далі, у той час як рухаючись від

(3, 4) за вертикаллю до (2, 4) і далі до (2.3), ми вертаємося через клітки (1, 3) і (1, 2) до вихідної клітки (3, 2) і утворюємо замкнутий цикл.

У побудованому ланцюжку, починаючи з клітки, що вводиться (яка вважається першою), позначають вершини: непарні - «+Θ», а парні «-Θ». Знаком «+» відзначають ті клітки, у яких обсяги перевезень мають збільшитися (такою, зокрема, є клітка, що вводиться до плану, оскільки вона має стати базисною). Знаком «-» - ті клітки, у яких перевезення зменшуються з метою збереження балансу. Серед множини кліток, позначених знаком «-», обирають клітку з найменшим значенням x_{ij}. Вона є кандидатом на вивід, тому що зменшення обсягу перевезень на більшу величину може призвести до від'ємних значень x_{ij} в інших «мінусових» клітках. Потім виконується перерахування плану за ланцюжком: до обсягів перевезень у клітках, позначених знаком «+», додається обсяг Θ, а з обсягів кліток, позначених знаком «-», він віднімається. У результаті введення однієї клітки й виведення іншої утворюється новий базисний план, для якого на наступній ітерації описані вище дії повторюються.

У нашому прикладі знаком «-» відзначені клітки (3, 4), (2, 3) і (1, 2), причому x_{34}=35, x_{23} =55, x_{12} =50. Обчисливши значення Θ= min{x_{34}, x_{23}, x_{12}} = 35, здійснюємо перетворення та переходимо до наступного базисного плану, показаному в таблиці 3.8.

Таблиця 3.8 - Оцінка оптимальності наступного базисного плану

		v_1=7	v_2=5	v_3=2	v_4=4	v_5=7	
		B_1	**B_2**	**B_3**	**B_4**	**B_5**	**Запаси**
u_1=0	**A_1**	7 30	5 15	2 80	8 -4	7 0	125
u_2=2	**A_2**	8 1	9 -2	4 20	6 40	9 0	60
u_3=-4	**A_3**	5 -2	1 35	9 -11	2 -2	3 80	115
	Заявки	30	50	100	40	80	300

Для знов отриманого плану повторюються дії стандартної ітерації: розраховуються потенціали та оцінки для небазисних кліток транспортної таблиці. Як можна бачити, план у таблиці 3.6 так само не є оптимальним (у клітці (2, 1) Δ_{21} =1>0), тому знову будуємо ланцюжок перетворення плану та переходимо до наступного базисного плану за ланцюжком в таблиці 3.9.

Таблиця 3.9 - Перетворення поточного базисного плану

		v_1=7	v_2=5	v_3=2	v_4=4	v_5=7	
		B_1	**B_2**	**B_3**	**B_4**	**B_5**	**Запаси**
u_1=0	**A_1**	7 30 -Θ	5 15	2 80 +Θ	8 -4	7 0	125
u_2=2	**A_2**	8 +Θ 1	9 -2	4 20 -Θ	6 40	9 0	60
u_3=-4	**A_3**	5 -2	1 35	9 -11	2 -2	3 80	115
	Заявки	30	50	100	40	80	300

Визначивши Θ=20, одержимо таблицю 3.10.

Таблиця 3.10 - Оптимальний базисний план

		$v_1=7$	$v_2=5$	$v_3=2$	$v_4=5$	$v_5=7$	
		B_1	**B_2**	**B_3**	**B_4**	**B_5**	**Запаси**
$u_1=0$	**A_1**	7 10	5 15	2 100	8 -3	7 0	125
$u_2=1$	**A_2**	8 20	9 -3	4 -1	6 40	9 -1	60
$u_3=-4$	**A_3**	5 -2	1 35	9 -11	2 -1	3 80	115
	Заявки	30	50	100	40	80	300

З транспортної таблиці 3.10 видно, що отриманий план є оптимальним, тому що всі оцінки для небазисних кліток $\Delta_{ij} \le 0$, тобто $u_i + v_j$ не перевищують відповідних цін c_{ij}. За даним планом обчислюється оптимальне (найменше) значення сумарних витрат на перевезення

$$L^*=10x_7+15x_5+100x_2+20x_8+40x_6+35x_1+80x_3=1020.$$

3.4 Випадок виродження

Якщо задача вироджена, то на якомусь етапі розв'язання може виявитися, що таблиця містить менше за $m+n$-1 заповнених кліток. Це, зокрема, може відбутися, якщо однакове мінімальне значення буде досягнуто відразу на кількох клітках, позначених знаком «-». Для подолання виродженості в транспортній задачі поточний план доповнюють необхідною кількістю нульових кліток (фіктивними перевезеннями) таким чином, щоб вони дозволяли розрахувати повну систему потенціалів, і далі діяти відповідно до правил описаного вище алгоритму. Тобто, якщо не вистачає k заповнених кліток, то їх вважають фіктивно заповненими. Фактично такий прийом є аналогом методу збурень для транспортної задачі як окремого випадку ЗЛП. До такого висновку легко прийти, якщо покласти, що фіктивні клітки, які додаються, містять певний малий обсяг ε.

Розглянемо приклад. З трьох кар'єрів до чотирьох склозаводів возять глину. Вартості перевезень, потужності кар'єрів і потреби заводів наведені в таблиці 3.11.

Таблиця 3.11 - Вихідні дані

	B_1	**B_2**	**B_3**	**B_4**	**Запаси**
A_1	3	9	7	4	50
A_2	6	8	10	6	65
A_3	5	4	7	6	50
Потреби	50	20	65	30	165

Опорний план визначимо за методом найменшої вартості (табл. 3.12).

Таблиця 3.12 - Початковий опорний план

		$v_1=3$	$v_2=5$	$v_3=8$	$v_4=4$	
		B_1	**B_2**	**B_3**	**B_4**	**Запаси**
$u_1=0$	**A_1**	3 50	9 -4	$+\Theta$ 7 1	$-\Theta$ 4 [0]	50
$u_2=2$	**A_2**	6 -1	8 -1	$-\Theta$ 10 35	$+\Theta$ 6 30	65
$u_3=-1$	**A_3**	5 -3	4 20	7 30	6 -3	50
	Потреби	50	20	65	30	165

У таблиці 3.12 заповнених кліток 5, а необхідно $m+n-1 = 3+4-1 = 6$. Потрібна одна фіктивно заповнена клітка. Будемо вважати клітку (1, 4) заповненою. Перевіримо план на оптимальність.

План не оптимальний, тому що $\Delta_{13} = 1$. $\Theta = \min(35{,}0) = 0$. План залишається таким же, але фіктивно заповненою буде клітка (1, 3). Перевіримо його на оптимальність.

Таблиця 3.13 - Перевірка базисного плану на оптимальність

		$v_1=3$	$v_2=4$	$v_3=7$	$v_4=3$	
		B_1	**B_2**	**B_3**	**B_4**	**Запаси**
$u_1=0$	**A_1**	3 50	9 -5	7 0	4 -1	50
$u_2=3$	**A_2**	6 0	8 -1	10 35	6 30	65
$u_3=0$	**A_3**	5 -2	4 20	7 30	6 -3	50
	Потреби	50	20	65	30	165

Всі $\Delta_{ij} \le 0$, отже план є оптимальним. Вартість перевезень при такому плані становить $L = 50*3+20*4+0*7+35*10+30*7+30*6 = 970$.

3.5 Транспортна задача за критерієм часу

Транспортна задача, окрім використання її з метою мінімізації витрат при транспортуванні вантажів, застосовується для оптимізації цілого ряду економічних процесів.

Природним застосуванням транспортної задачі є **задача мінімізації часу транспортування**. Наприклад, на підприємстві є задана кількість виробничих ліній, що працюють із заданою продуктивністю. Є також задана кількість складів продукції. Відомий час транспортування від кожної лінії до кожного складу. Задача зводиться до вибору таких маршрутів, при використанні яких час, затрачуваний на транспортування, буде мінімальним.

Іншою задачею є вибір оптимального варіанта з використання наявного встаткування. Задача полягає у виборі оптимального плану виробництва виробів при мінімізації часу, необхідного на їхнє виготовлення. Ця задача може бути сформульована в такий спосіб. Є задана кількість видів устаткування, на кожному з яких може бути виготовлена деяка задана кількість виробів. Відома продуктивність для кожного типу встаткування й кожного виробу. Необхідно скласти план, при якому підприємство затратить мінімальну кількість часу на виготовлення всіх виробів.

Розглянемо приклад. Визначити мінімальний час виготовлення трьох видів виробів на трьох типах устаткування. Продуктивність устаткування (a_i) і план виробництва виробів (b_j) наведені в таблиці 3.14.

Таблиця 3.14 - Вихідні дані

	B_1	B_2	B_3	План
A_1	15	7	8	240
A_2	9	4	11	80
A_3	6	3	7	180
Продуктивність	200	160	140	500

Задача є збалансованою. Знайдемо початковий базисний план, використовуючи метод мінімальної вартості. Найменша продуктивність, що дорівнює 3, записана в клітці (3, 2). Потужність устаткування третього типу дорівнює 180, а необхідна кількість виробів другого типу дорівнює 160, тобто min{180,160}=160. Тому в клітку (3, 2) уписуємо число 160. Другий стовпець випадає з подальшого розгляду, тому що продуктивність устаткування другого типу вичерпана. Аналогічно заповнюємо інші клітки таблиці 3.15.

Таблиця 3.15 - Визначення початкового опорного плану

		$v_1=15$	$v_2=12$	$v_3=8$	
		B_1	B_2	B_3	План
$u_1=0$	A_1	15 -Θ 100	7 +Θ 5	8 140	240
$u_2=-6$	A_2	9 80	4 2	11 -9	80
$u_3=-9$	A_3	6 +Θ 20	3 -Θ 160	7 -8	180
	Продуктивність	200	160	140	500

Відповідно до отриманого плану час виготовлення всіх виробів становить L = 100*15+140*8+80*9+20*6+160*3 = 3940. Визначимо потенціали u_i і v_j для зайнятих кліток: $u_1=0$; $v_1=15$; $v_3=8$; $u_2=9-15=-6$; $u_3=6-15=-9$; $v_2=3-(-9)=12$. Обчислимо оцінки для вільних кліток: $\Delta_{12}=(12+0)-7=5$; $\Delta_{22}=(12-6)-4=2$; $\Delta_{23}=(-6+8)-11=-9$; $\Delta_{33}=(-9+8)-7=-8$.

Оскільки оцінки вільних кліток (1, 2) і (2, 2) додатні, план не оптимальний. Перейдемо до нового базисного плану, побудувавши цикл для клітки (1, 2), $\Theta=\min\{100, 160\}=100$. Новий базисний план наведений у табл. 3.16. Із клітки (1,1) вироби в кількості 100 од. переносимо до клітки (1,2). Із клітки (3,2) вироби в кількості 100 од. переносимо до клітки (3,1).

Таблиця 3.16 - Визначення наступного базисного плану

		B₁ (v₁=10)	B₂ (v₂=7)	B₃ (v₃=8)	План
$u_1=0$	**A_1**	15 -5	7 100	8 140	240
$u_2=-1$	**A_2**	-Θ 9 80	+Θ 4 2	11 -4	80
$u_3=-4$	**A_3**	+Θ 6 120	-Θ 3 60	7 -3	180
	Продуктивність	200	160	140	500

Відповідно до отриманого плану час виготовлення всіх виробів складе L=100*7+140*8+80*9+120*6+60*3 = 3440. Перевірка плану на оптимальність показує, що він не оптимальний, тому що в клітці (2,2) Δ_{22} додатна. Побудуємо цикл для цієї клітки й перетворимо план при Θ=60. Новий базисний план наведений у таблиці 3.17.

Таблиця 3.17 - Визначення наступного базисного плану

		B₁ (v₁=12)	B₂ (v₂=7)	B₃ (v₃=8)	План
$u_1=0$	**A_1**	15 -3	7 100	8 140	240
$u_2=-3$	**A_2**	9 20	4 60	11 -6	80
$u_3=-6$	**A_3**	6 180	3 -2	7 -5	180
	Продуктивність	200	160	140	500

Перевірка показує оптимальність плану. Відповідно до отриманого плану час виготовлення всіх виробів складе L = 100*7+140*8+20*9+60*4+180*6 = =3320, і він є мінімальним.

Контрольні запитання

1. Які специфічні властивості дозволяють виділити транспортні задачі з множини задач лінійного програмування в окремий клас?
2. Опишіть методи побудови припустимого плану транспортної задачі.
3. Скільки ненульових елементів повинен містити невироджений базисний план транспортної задачі?
4. Сформулюйте критерій оптимальності для припустимого плану транспортної задачі.
5. На чому заснований метод потенціалів?
6. З чого випливає критерій оптимальності припустимого плану транспортної задачі?
7. Перелічіть основні етапи методу потенціалів.
8. Які умови мають бути дотримані при побудові ланцюжка перетворення плану в методі потенціалів?
9. Що потрібно робити при виникненні ситуації виродженості поточного плану в транспортній задачі?

ЗМ 2 Економічна інтерпретація і аналіз оптимальних планів лінійних оптимізаційних моделей

ТЕМА 4
ТЕОРІЯ ПОДВІЙНОСТІ ТА ДВОЇСТІ ОЦІНКИ В АНАЛІЗІ РОЗВ'ЯЗКІВ ЛІНІЙНИХ ОПТИМІЗАЦІЙНИХ МОДЕЛЕЙ

4.1 Пряма та двоїста задачі як пара сполучених задач

Більшість лінійних оптимізаційних задач можна розглядати як економічні задачі про розподіл ресурсів $b_1, b_2, ..., b_m$ між різними споживачами, зокрема, між технологічними процесами $\overline{A}_1, \overline{A}_2, ..., \overline{A}_n$.

Будь-який припустимий розв'язок лінійної задачі $x_1, x_2, ..., x_n$ дає конкретний розподіл ресурсів, що вказує частку кожного з них, яка буде використаною при реалізації певного технологічного процесу.

Загальна задача лінійного програмування формулюється в такий спосіб:

знайти

$$L = c_1x_1 + c_2x_2 + ... + c_nx_n \to \max, \tag{4.1}$$

де x задовольняє обмеженням

$$\begin{aligned} &a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n \le b_1, \\ &a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n \le b_2, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n \le b_m, \\ &\qquad x_j \ge 0. \end{aligned} \tag{4.2}$$

Тут a_{ij} – витрати i-го ресурсу на j-й технологічний процес. Зміст обмежень – сумарні витрати i-го ресурсу повинні не перевищувати кількість цього ресурсу, наявного в нашому розпорядженні b_i.

Для будь-якого технологічного процесу актуальним є так само запитання про те, щоб витрати ресурсів були мінімальними. Позначимо u_i ціну одиниці i-го ресурсу. Тоді можна записати вимогу, щоб вартість ресурсів була мінімальною

$$L' = b_1u_1 + b_2u_2 + ... + b_mu_m \to \min,$$

а величина витрат на кожний вид технологічного процесу не повинна бути меншою за його вартість

$$\begin{aligned} &a_{11}u_1 + a_{12}u_2 + ... + a_{1m}u_m \ge c_1, \\ &a_{12}u_1 + a_{22}u_2 + ... + a_{m2}u_m \ge c_2, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \; . \\ &a_{1n}u_1 + a_{2n}u_2 + ... + a_{mn}u_m \ge c_n, \\ &\qquad u_i \ge 0. \end{aligned}$$

Отже, одержали задачу:

знайти такий розв’язок u*, що перетворює на мінімум цільову функцію

$$L' = \sum_{i=1}^{m} b_i u_i \to \min \qquad (4.3)$$

і задовольняє обмеженням

$$\begin{array}{l} a_{11}u_1 + a_{12}u_2 + ... + a_{1m}u_m \geq c_1, \\ a_{12}u_1 + a_{22}u_2 + ... + a_{m2}u_m \geq c_2, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots \ . \\ a_{1n}u_1 + a_{2n}u_2 + ... + a_{mn}u_m \geq c_n, \\ \qquad u_i \geq 0. \end{array} \qquad (4.4)$$

Дану задача називають двоїстою стосовно вихідної задачі, що називається прямою задачею.

Порівняємо умови прямої та двоїстої задач у матричній формі:

Пряма	Двоїста
Знайти таке $\overline{X}$*, що	Знайти таке $\overline{U}$*, що
$\overline{C}^T\,\overline{X}^* \to \max$	$\overline{B}^T\,\overline{U}^* \to \min$
при обмеженнях	при обмеженнях
$\overline{A}\overline{X} \leq \overline{B}, \quad \overline{X} \geq 0$.	$\overline{A}^T\overline{U} \geq \overline{C}, \quad \overline{U} \geq 0$.

Загальна схема побудови двоїстої задачі. Якщо задано загальну задачу ЛП, де множина припустимих планів D визначається системою рівнянь і нерівностей (4.2), то двоїстою стосовно неї називається загальна задача ЛП, де D' визначається системою рівнянь і нерівностей (4.4).

При переході від прямої задачі ЛП до двоїстої:

- тип оптимуму змінюється на протилежний, тобто якщо пряма задача є задачею максимізації, то двоїста буде задачею мінімізації та навпаки;

- вектор коефіцієнтів цільової функції $\overline{C}$ та стовпець обмежень $\overline{B}$ міняються місцями. Тобто коефіцієнти цільової функції прямої задачі c_1, c_2, ..., c_n стають вільними членами обмежень двоїстої задачі, а вільні члени обмежень прямої задачі b_1, b_2, ..., b_m стають коефіцієнтами цільової функції двоїстої задачі;

- матрицю обмежень двоїстої задачі одержують транспонуванням матриці обмежень прямої задачі $\overline{A}$;

- множину індексів змінних, на які накладено умову невід’ємності в прямій задачі (наприклад, $x_j \geq 0$ або $u_i \geq 0$), визначають номера обмежень, що мають форму нерівностей у двоїстій задачі ($a_j u \geq c_j$ або $a_i x \leq b_i$);

- множину номерів обмежень, що мають форму нерівностей у прямій задачі (наприклад, $a_i x \leq b_i$ або $a_j u \geq c_j$), визначає множина індексів змінних, на які накладено умову невід’ємності, у двоїстій задачі ($u_i \geq 0$ або $x_j \geq 0$).

З наведених властивостей пари задач випливає важлива властивість - **симетричність відносини подвійності**, тобто задача, що є двоїстою стосовно двоїстої, збігається з прямою (вихідною) задачею.

Тим самим має смисл говорити про пару взаємно двоїстих задач.

4.2 Основні теореми подвійності, їхній економічний зміст

Фундаментальні властивості, якими володіють двоїсті задачі лінійного програмування, можна сформулювати у вигляді тверджень, що приводяться нижче. Їх зазвичай називають теоремами подвійності.

Теорема 4.1 (перша теорема подвійності). Якщо $\overline{X}_0$ та $\overline{U}_0$ - припустимі плани для пари двоїстих задач, тобто якщо

$$\overline{A}\ \overline{X}_0 \le \overline{B} \qquad \text{і} \qquad \overline{A}^{\mathrm{T}}\ \overline{U}_0 \ge \overline{C},$$

то

$$\overline{C}^T\ \overline{X}_0 \le \overline{B}^T\ \overline{U}_0,$$

тобто значення цільової функції прямої задачі ніколи не перевищують значень цільової функції двоїстої задачі.

Доказ.

Оскільки $\overline{U}_0$ - припустимий план, то

$$\overline{A}^{\mathrm{T}}\ \overline{U}_0 \ge \overline{C}; \tag{4.5}$$

оскільки $\overline{X}_0$ - припустимий план, то

$$\overline{A}\ \overline{X}_0 \le \overline{B}. \tag{4.6}$$

Помножимо (4.5) на $\overline{X}_0^T$

$$\overline{X}_0^T \overline{A}^T \overline{U}_0 \ge \overline{X}_0^T \overline{C}; \tag{4.7}$$

помножимо (4.6) на $\overline{U}_0^T$

$$\overline{X}_0 \overline{A} \overline{U}_0^T \le \overline{U}_0^T \overline{B} \tag{4.8}$$

і порівняємо (4.7) з (4.8). Оскільки

$$\overline{X}_0 \overline{A} \overline{U}_0^T = \left(\overline{X}_0^T \overline{A}^T \overline{U}_0\right)^T, \qquad \text{то } \overline{U}_0^T \overline{B} \ge \overline{X}_0^T \overline{C},$$

або

$$\overline{U}_0 \overline{B}^T \ge \overline{X}_0 \overline{C}^T.$$

Зауваження. Теорема 4.1, зрозуміло, вірна й для оптимальних планів взаємно двоїстих задач: $\overline{U}^* \overline{B}^T \ge \overline{X}^* \overline{C}^T$, де $\overline{X}^*$ і $\overline{U}^*$ - будь-які оптимальні плани задач. Насправді, як буде видно з подальшого, справедливою є рівність $\overline{U}^* \overline{B}^T = \overline{X}^* \overline{C}^T$.

Теорема 4.2 (друга теорема подвійності). Якщо для деяких припустимих планів $\overline{X}_0$ і $\overline{U}_0$ взаємно двоїстих задач виконується рівність

$$\overline{C}^{\mathrm{T}}\ \overline{X}_0 = \overline{B}^{\mathrm{T}}\ \overline{U}_0,$$

то $\overline{X}_0$ і $\overline{U}_0$ є оптимальними планами цих задач.

Доказ.

Відповідно до теореми 4 для всіх припустимих розв’язків $\overline{X}$ і $\overline{U}$ справедлива нерівність

$$\overline{C}^T \overline{X} \le \overline{B}^T \overline{U},$$

але оскільки з умови теореми

$$\overline{C}^T \overline{X}_0 = \overline{B}^T \overline{U}_0,$$

то $\overline{C}^T\overline{X}_0$ – найбільше з можливих значень цільової функції, тобто

$$\overline{C}^T\overline{X} \le \overline{C}^T\overline{X}_0,$$

цільова функція від $\overline{X}_0$ – найбільша, таким чином, $\overline{X}_0$ є оптимальним значенням.

Аналогічно для цільової функції двоїстої задачі

$$\overline{B}^T\overline{U}_0 \le \overline{B}^T\overline{U},$$

тобто $\overline{B}^T\overline{U}_0$ – найменше з можливих значень цільової функції двоїстої задачі, отже $\overline{U}_0$ – оптимальне значення.

Отже, якщо для деяких припустимих планів $\overline{X}_0$ і $\overline{U}_0$ прямої та двоїстої задач їхні цільові функції дорівнюють одна одній, то $\overline{X}_0$ та $\overline{U}_0$ – оптимальні плани пари сполучених задач.

Теорема 4.3. Якщо цільова функція в прямій задачі не обмежена зверху, то двоїста до неї задача не має припустимих планів.

Доказ.

Якщо припустити, що у двоїстій задачі існує хоча б один припустимий план $\overline{U}_0$, то відповідно до теореми 4.2, для будь-якого припустимого плану $\overline{X}_0$ прямої задачі справедлива нерівність $\overline{C}^T\overline{X} \le \overline{B}^T\overline{U} < +\infty$. Останнє означає, що цільова функція прямої задачі обмежена зверху. Оскільки це суперечить умові теореми, припущення про існування припустимих планів двоїстої задачі невірне.

Наступне твердження, що відоме як теорема рівноваги, використовується при перевірці оптимальності планів ЗЛП.

Теорема 4.4 Нехай $\overline{X}*$ і $\overline{U}*$ - оптимальні плани канонічної та двоїстої стосовно неї задач. Якщо j-та компонента плану $\overline{X}*$ строго додатна ($x^*_j>0$), то відповідне j-е обмеження двоїстої задачі виконується як рівність: $a_{1j}u_1+\ldots+a_{mj}u_m=c_j$; якщо j-та компонента плану $\overline{X}*$ має нульове значення ($x^*_j = 0$), то j-е обмеження двоїстої задачі виконується як нерівність: $a_{1j}u_1+\ldots+a_{mj}u_m>c_j$.

Доказ.

Вектори $\overline{X}*$ і $\overline{U}*$, як припустимі плани, задовольняють обмеженням

$\overline{A}\overline{X}*=\overline{B}$, $\overline{X}*>0$ - прямої задачі

$\overline{A}^T\overline{U}*-\overline{C}^T \ge 0$, $\overline{U}*>0$ - двоїстої задачі.

Запишемо скалярний добуток

$$\left(\overline{A}^T\overline{U}*-\overline{C}^T\right)\overline{X}*=\overline{A}^T\overline{X}*\overline{U}*-\overline{C}^T\overline{X}*=\overline{B}^T\overline{U}*-\overline{C}^T\overline{X}*.$$

Одержали різницю цільових функцій прямої і двоїстої задач. На підставі другої теореми подвійності оптимальні значення цільових функцій взаємно двоїстих задач збігаються. Отже скалярний добуток

$$\left(\overline{A}^T \overline{U}^* - \overline{C}^T\right)\overline{X}^* = 0 .$$

Але скалярний добуток двох невід'ємних векторів може дорівнювати нулю тільки в тому випадку, якщо всі попарні добутки їхніх відповідних координат дорівнюють нулю. Тоді, якщо $x_j > 0$, то $\sum_i a_{ij}u_i - c_j = 0$ або $\sum_i a_{ij}u_i = c_j$. А якщо $x_j = 0$, то можливо, що $\sum_i a_{ij}u_i - c_j > 0$ або $\sum_i a_{ij}u_i > c_j$.

Практичне значення теорем подвійності полягає в тому, що вони дозволяють замінити процес розв'язання основної задачі на розв'язання двоїстої, яке в певних випадках може виявитися простішим. Наприклад, задачу, область припустимих значень якої описується двома рівняннями, що зв'язують шість змінних (m = 2, n = 6), не можна вирішити графічним методом. Проте даний метод можна застосувати для розв'язання двоїстої до неї задачі, що має тільки дві змінні.

Оптимальний розв'язок двоїстої задачі можна одержати з таблиці, отриманої на фінальній ітерації процедури симплекс-методу. Елементи індексного рядка цієї таблиці L_j обчислюють відповідно до виразу (2.9) $L_j = \sum_{i=1}^{m} c_j a_{ij}$, де c_j – елементи вектора-рядка, що містить коефіцієнти цільової функції прямої задачі при змінних, що є базисними в оптимальному плані; a_{ij} - елементи матриці $\overline{D}^{-1}$, що зворотна до матриці $\overline{D}$. Матриця $\overline{D}$ складається з базисних векторів оптимального плану, компоненти яких узяті з початкового опорного плану задачі. Оптимальний план двоїстої задачі визначається співвідношенням

$$\overline{U}^* = \overline{c}_{j\text{баз}}\overline{D}^{-1} . \tag{4.9}$$

Зворотна матриця $\overline{D}^{-1}$ завжди міститься в останній симплекс-таблиці в тих стовпчиках, де в першій таблиці задачі знаходилась одинична матриця. (Нагадаємо, що добуток матриці на її зворотну дає одиничну матрицю, в якої діагональні елементи дорівнюють 1, а всі інші дорівнюють 0).

Отже, зв'язок між оптимальними розв'язками прямої та двоїстої задач і елементами індексних рядків L_j симплекс-таблиць, що відповідають цим розв'язкам, виражається наступними співвідношеннями:

$$a_{0,n+i}^{ПР} = u_i^*, \quad i = \overline{1,m},$$

$$-a_{0,m+j}^{ДВ} = x_j^*, \quad j = \overline{1,n},$$

де n – кількість змінних прямої задачі;

m – кількість обмежень прямої задачі;

$a_{0,n+i}^{ПР}$ - ($n+i$)-й елемент індексного рядка симплекс-таблиці прямої задачі, що містить оптимальний план;

$a_{0,m+j}^{ДВ}$ - ($m+j$)-й елемент індексного рядка симплекс-таблиці двоїстої задачі, що містить оптимальний план.

4.3 Двоїсті оцінки та дефіцитність ресурсів

У різних джерелах компоненти оптимального плану двоїстої задачі називають **двоїстими оцінками** або **тіньовими цінами**.

На підставі теорем подвійності для пари задач ЛП у загальній формі можна сформулювати певні важливі з погляду економічної інтерпретації слідства. Зокрема, з теореми 4.4 випливає, що якщо при реалізації оптимального плану прямої задачі i-е обмеження виконується як строга нерівність, то оптимальне значення відповідної двоїстої змінної дорівнює нулю:

$$a_{i1}x_1^* + \cdots + a_{in}x_n^* < b_i \text{, то } u_i^* = 0 .$$

Це означає, що якщо певний ресурс b_i є в надлишковій кількості (не використовується повністю при реалізації оптимального плану), то i-е обмеження стає неістотним і тіньова оцінка такого ресурсу дорівнює нулю. Отже, тіньові оцінки характеризують **дефіцитність ресурсів**.

Якщо при реалізації оптимального плану двоїстої задачі j-е обмеження виконується як строга нерівність, то оптимальне значення відповідної змінної в оптимальному плані прямої задачі має дорівнювати нулю

$$a_{1j}u_1^* + \cdots + a_{mj}u_m^* > c_j \text{, то } x_j^* = 0 .$$

Даний факт виражає **принцип рентабельності виробництва**. З огляду на економічний зміст двоїстих оцінок $u_1^*, \cdots, u_m^*$, вираз $a_{1j}u_1^* + \cdots + a_{mj}u_m^*$ можна інтерпретувати як питомі витрати на j-й технологічний процес. Отже, якщо ці витрати перевищують прибуток від реалізації одиниці j-го продукту, виробництво j-го продукту є нерентабельним і не повинне бути присутнім в оптимальному виробничому плані ($x_j^* = 0$).

Незважаючи на можливі аналогії, які можуть виникнути у зв'язку з такими фундаментальними поняттями економічної теорії, як граничні витрати та граничний дохід, двоїсті оцінки не слід однозначно ототожнювати з цінами (хоча такі спроби іноді вживали на початковій стадії становлення дослідження операцій як науки).

Контрольні запитання

1. Поясніть суть подвійності в лінійному програмуванні.
2. Складіть просту лінійну оптимізаційну модель і запишіть до неї двоїсту. Дайте економічну інтерпретацію двоїстих оцінок.
3. Скільки змінних і скільки обмежень має двоїста задача стосовно прямої задачі?
4. Поясніть економічний зміст першої теореми подвійності.
5. Поясніть економічний зміст другої теореми подвійності.

6. У чому полягає економічний зміст третьої теореми подвійності?
7. Сформулюйте правила побудови двоїстих задач.
8. Як, маючи оптимальний розв'язок прямої задачі, можна одержати оптимальний розв'язок двоїстої задачі?

ТЕМА 5
АНАЛІЗ ЛІНІЙНИХ МОДЕЛЕЙ ОПТИМІЗАЦІЙНИХ ЗАДАЧ

5.1 Аналіз розв'язків лінійних оптимізаційних моделей

Традиційна економічна інтерпретація двоїстої задачі ЛП базується на моделі найпростішої задачі виробничого планування. У ній кожний j-й елемент вектора $\overline{X}$ розглядається як план випуску продукції певного виду в натуральних одиницях, c_j - ціна одиниці продукції j-го виду, $\overline{A}_j$ - вектор, що визначає технологію витрати наявних m ресурсів на виробництво одиниці продукції j-го виду, $\overline{B}$ - вектор обмежень на обсяги цих ресурсів.

Припустимо, що для певних значень $\overline{A}_j$, $\overline{B}$ і c_j знайдений оптимальний план x^* прямої задачі, що максимізує сумарний дохід $\max\limits_{x \in D}\{cx\} = cx^*$, і визначені оптимальні оцінки сировини, тобто оптимальний план двоїстої задачі u^*. З виразу цільової функції двоїстої задачі $L'^* = b_1 u_1^* + b_2 u_2^* + \ldots + b_m u_m^*$ видно, що величина двоїстої оцінки u_i^* показує, наскільки зросте максимальне значення цільової функції прямої задачі при збільшенні кількості сировини відповідного виду на одну одиницю. Отже, змінні двоїстої задачі $u_1^*, \cdots, u_m^*$ за своїм змістом є оцінками потенційної можливості одержання додаткового прибутку за рахунок збільшення обсягу відповідного ресурсу в умовах оптимального функціонування керованого економічного об'єкта.

Виникає питання про те, як змінюватиметься оптимальний план x^* при зміні компонентів вектору обмежень $\overline{B}$ і, зокрема, при яких варіаціях $\overline{B}$ оптимальний план x^* залишиться оптимальним. Ця задача одержала назву проблеми **стійкості оптимального плану**. Очевидно, що дослідження стійкості x^* має й безпосереднє практичне значення, тому що в реальному виробництві обсяги доступних ресурсів b_i можуть істотно коливатися після ухвалення планового розв'язку x^*.

Коли вектор обмежень $\overline{B}$ отримує збільшення Δb, виникають відповідні варіації для оптимального плану прямої задачі $x^*(b+\Delta b)$ і значення цільової функції $L[x^*(b+\Delta b)]$. Припустимо, що збільшення Δb таке, що воно не призводить до зміни оптимального базису задачі, тобто $x^*(b+\Delta b) \geq 0$. Уведемо функцію $F(b)$, що повертає оптимальне значення цільової функції задачі для різних значень вектора обмежень $\overline{B}$

$$F(b) = \max_{x \in D(b)}\{cx\}. \qquad (5.1)$$

Розглянемо відношення її приросту $F(b+\Delta b)$-$F(b)$ до приросту аргументу Δb. Якщо для певного і спрямувати $\Delta b \to 0$, одержимо

$$\lim_{\Delta b_i \to 0} \frac{F(b+\Delta b)-F(b)}{\Delta b_i} = \frac{\partial F(b)}{\partial b_i}. \qquad (5.2)$$

З огляду на те, що відповідно до теореми 4.2 цільові функції пари сполучених задач при оптимальних планах дорівнюють одна одній, запишемо

$$F(b) = \sum_{j=1}^{n} c_j x_j^* = \sum_{i=1}^{m} b_i u_i^* . \qquad (5.3)$$

Підставимо (5.3) до (5.2) і одержимо вираз

$$\frac{\partial F(b)}{\partial b_i} = \frac{\partial \left(\sum_{i=1}^{m} b_i u_i^* \right)}{\partial b_i} = u_i^* . \qquad (5.4)$$

Теорема 5.1 В оптимальному плані двоїстої задачі значення змінної u_i* чисельно дорівнює частинній похідній цільової функції *L’** за відповідним аргументом b_i.

Звідси випливає економічна інтерпретація оптимальних змінних двоїстої задачі:

Кожний елемент $u_1^*, \cdots, u_m^*$ може розглядатися як гранична (миттєва) оцінка внеску і-го ресурсу в сумарний дохід *L** при оптимальному розв’язку $x_1^*, \cdots, x_n^*$.

Інакше кажучи, u_i^* дорівнює приросту доходу, що виникає при збільшенні ресурсу і на одиницю за умови оптимального використання ресурсів.

Отже, якщо знайдено оптимальний план прямої задачі, можна провести аналіз стійкості двоїстих оцінок щодо змін компонентів вектора $\overline{B}$. Це дозволяє оцінити стабільність оптимального плану двоїстої задачі щодо зміни обмежень прямої задачі та ступінь впливу зміни $\overline{B}$ на максимальне значення цільової функції, а також визначити найбільш доцільний варіант можливих змін $\overline{B}$.

План двоїстої задачі не змінюється для всіх значень $b_i + \Delta b_i$, при яких стовпець вектора $\overline{B}*$ останньої симплекс-таблиці не містить від’ємних чисел, тобто коли серед компонентів вектора

$$\overline{B}* = \begin{vmatrix} b_1 + \Delta b_1 \\ b_2 + \Delta b_2 \\ \cdots\cdots\cdots \\ b_m + \Delta b_m \end{vmatrix} \qquad (5.5)$$

немає від’ємних. $\overline{B}*$ - матриця, складена з компонентів векторів базису, що визначає оптимальний план задачі, тому що базисні компоненти оптимального плану знаходяться у стовпці вектора $\overline{B}$ останньої симплекс-таблиці.

Елементи (n+і)-го стовпця a_{ij} останньої симплекс-таблиці, що містить оптимальну оцінку і-го ресурсу u_i*, показують, на скільки одиниць зміняться компоненти оптимального плану x* при збільшенні обсягу цього ресурсу на одиницю, тобто $x_j^*(b_i+\Delta b_i)=x_j^*(b_i)+a_{i,n+i}\Delta b_i$. Припустимі інтервали зміни для i-го ресурсу можна визначити з умови:

$$\overline{B}^* = \left| \begin{array}{l} x_1 + a_{1,n+i}\Delta b_i \geq 0 \\ x_2 + a_{2,n+i}\Delta b_i \geq 0 \\ \dots\dots\dots\dots\dots\dots \\ x_n + a_{n,n+i}\Delta b_i \geq 0 \end{array} \right| . \qquad (5.6)$$

Розглянемо приклад. Нехай таблиця 5.1 є останньою симплекс-таблицею, що містить оптимальний план.

Таблиця 5.1 – Оптимальний план прямої задачі

Базис	$Cj_{баз}$	***Cj***	6	5	0	0	0
		B	A_1	A_2	A_3	A_4	A_5
A_1	6	83	1	0	4/7	-3/7	0
A_2	5	36	0	1	-3/7	4/7	0
A_5	0	21	0	0	11/28	-11/7	1
індексний рядок	L_j	678	6	5	9/7	2/7	0
	Δ_j		0	0	9/7	2/7	0

Звідки оптимальний план прямої задачі $x^*=(83; 36; 0; 0; 21)$, оптимальний план двоїстої задачі $u^*=(\frac{9}{7}; \frac{2}{7}; 0)$. Більш дефіцитною є сировина S_1, оскільки її тіньова оцінка вище й відповідно сильніше вплив на величину цільової функції.

Очевидно, що збільшення доходу можна отримати тільки шляхом зміни оптимального плану прямої задачі. З таблиці видно, що при збільшенні на 1 одиницю кількості сировини S_1, дохід збільшиться на $\frac{9}{7}$ грн. Це відбудеться якщо виробництво виробів А збільшити на $\frac{4}{7}$ одиниці, а виробництво виробів Б знизити на $\frac{3}{7}$ одиниці, при цьому витрата сировини S_3 зросте на $\frac{11}{28}$ одиниці. Новий оптимальний план прямої задачі матиме вигляд $x^{*\prime}=(83\frac{4}{7}; 35\frac{4}{7}; 0; 0; 20\frac{17}{28})$, а прибуток становитиме $L^*=6*83\frac{4}{7}+5*35\frac{4}{7}+0*0+ +0*20\frac{17}{28} = 679{,}286$ грн.

Визначимо інтервали стійкості двоїстих оцінок. Для ресурсу 1 відповідно до елементів стовпчика A_3 маємо

$$x^* = (83 + 0{,}57\Delta b_1; 36 - 0{,}43\Delta b_1; 0; 0; 21 + 0{,}39\Delta b_1).$$

Запишемо вектор $\overline{B}^*$ з умовами його невід'ємності та визначимо межі припустимих значень Δb_1

$$\overline{B}^* = \left| \begin{array}{l} 83+0{,}57\Delta b_1 \geq 0 \\ 36-0{,}43\Delta b_1 \geq 0 \\ 21+0{,}39\Delta b_1 \geq 0 \end{array} \right| = \left| \begin{array}{l} \Delta b_1 \geq -66{,}7 \\ \Delta b_1 \leq 83{,}7 \\ \Delta b_1 \geq -53{,}8 \end{array} \right|$$

Оптимальний план двоїстої задачі залишиться незмінним, якщо Δb_1 належить інтервалу $-53{,}8 \le \Delta b_1 \le 83{,}7$, а перший ресурс $440-53{,}7 \le b_1 \le 440+83{,}7$ або $386{,}3 \le b_1 \le 523{,}7$.

Аналогічно для ресурсу 2 відповідно до елементів стовпчика A_4 запишемо

$$x^* = (83-0{,}43\Delta b_2;\ 36+0{,}57\Delta b_2;\ 0;\ 0;\ 21-1{,}57\Delta b_2).$$

$$\overline{B}^* = \begin{vmatrix} 83\text{-}0{,}43\Delta b_2 \ge 0 \\ 36\text{+}0{,}57\Delta b_2 \ge 0 \\ 21\text{-}1{,}57\Delta b_2 \ge 0 \end{vmatrix} = \begin{vmatrix} \Delta b_2 \le 193 \\ \Delta b_2 \ge \text{-}63{,}2 \\ \Delta b_2 \le 13{,}4 \end{vmatrix}$$

Оптимальний план двоїстої задачі залишиться незмінним, якщо Δb_2 належить інтервалу $-63{,}2 \le \Delta b_2 \le 13{,}4$, а другий ресурс $393-63{,}2 \le b_2 \le 393+13{,}4$ або $329{,}8 \le b_2 \le 406{,}4$.

Таким чином, якщо збільшення кількості ресурсів S_1 належить проміжку $-53{,}8<\Delta b_1<83{,}7$, а кількість інших ресурсів незмінна, або збільшення кількості ресурсів S_2 належить проміжку $-63{,}2<\Delta b_2<83{,}7$, а кількість інших ресурсів незмінна, то двоїста задача має той самий оптимальний план u^* = (1,286; 0,286; 0).

Стосовно прямої задачі, можна показати, що при зміні кількості першого ресурсу S_1 у межах $386{,}3 \le b_1 \le 523{,}7$ можливий дохід підприємства лежить у межах $609{,}7 \le L^* \le 784{,}3$ а оптимальний план прямої задачі

$$(52{,}3;\ 59{,}1;\ 0;\ 0;\ 0{,}02) \le x^* \le (131;\ 0;\ 0;\ 0;\ 53{,}6).$$

При зміні кількості другого ресурсу S_2 у межах $329{,}8 \le b_2 \le 406{,}4$ можливий дохід підприємства лежить у межах $661 \le L^* \le 681{,}6$ а оптимальний план прямої задачі $(110;\ 0;\ 0;\ 0;\ 53{,}6) \le x^* \le (77;\ 43{,}6;\ 0;\ 0;\ 0)$.

Розраховані інтервали стосуються випадків зміни тільки одного ресурсу. У випадку одночасної зміни обсягу всіх або кількох ресурсів для визначення нового оптимального плану використовують одно з основних співвідношень обчислювальної процедури симплекс-методу:

$$x^* = \overline{D}^{-1} * \overline{B}, \tag{5.7}$$

де $\overline{D}$ - матриця, що складається з базисних векторів оптимального плану, компоненти яких узяті з початкового опорного плану;

$\overline{B}$ - вектор обмежень.

У розглянутому числовому прикладі матриці $\overline{D}$ та $\overline{D}^{-1}$ відповідно мають вигляд

$$\overline{D} = \begin{vmatrix} 4 & 3 & 0 \\ 3 & 4 & 0 \\ 3 & 5 & 1 \end{vmatrix}; \qquad \overline{D}^{-1} = \begin{vmatrix} 4/7 & -3/7 & 0 \\ -3/7 & 4/7 & 0 \\ 11/28 & -11/7 & 1 \end{vmatrix}$$

Нехай новий вектор обмежень

$$\overline{B} = \begin{vmatrix} 440+84 \\ 393+13{,}4 \\ 450+0 \end{vmatrix} = \begin{vmatrix} 524 \\ 406{,}4 \\ 450 \end{vmatrix};$$

тоді новий оптимальний план визначиться в такий спосіб:

$$x^* = \begin{vmatrix} 4/7 & -3/7 & 0 \\ -3/7 & 4/7 & 0 \\ 11/28 & -11/7 & 1 \end{vmatrix} * \begin{vmatrix} 524 \\ 406,4 \\ 450 \end{vmatrix} = \begin{vmatrix} 125,26 \\ 7,66 \\ 17,2 \end{vmatrix}$$

Тобто x^*=(125,26; 7,66; 0; 0; 17,2), при якому прибуток дорівнюватиме 790 грош. од.

5.2 Аналіз параметричної стійкості розв'язків ЗЛП

З погляду економічної інтерпретації задача дослідження параметричної стійкості може бути розглянута як вивчення тих меж коливання цін на продукцію керованого підприємства (фірми), при яких прийнятий план випуску продукції продовжує залишатися оптимальним. Таким чином, питання стійкості оптимального плану ЗЛП може бути поставленим для випадку варіації коефіцієнтів цільової функції

$$c_j, \quad j = \overline{1,n}.$$

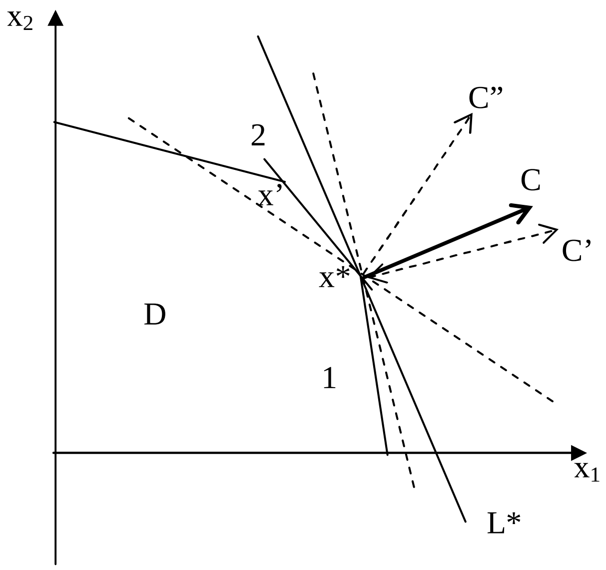

Рисунок 5.1 - Графічна інтерпретація параметричної стійкості оптимального плану

Зміст проблеми стійкості оптимального плану ЗЛП стосовно варіацій цільової функції можна проілюструвати за допомогою першої геометричної інтерпретації. На рисунку 5.1 зображена множина припустимих планів *D* певної задачі ЛП. Як видно з рисунку, цільова функція *L* досягає екстремального значення в точці x^*, а зміні її коефіцієнтів від c до c' або c'' на рисунку відповідає поворот лінії рівня відносно x^*. Активним обмеженням (тобто таким, що обертаються на рівність) у точці x^* відповідають лінії 1 і 2. Доти, поки при повороті, викликаному зміною вектора c, лінія рівня цільової функції не виходить за межі утвореної лініями обмежень множини, x^* залишається оптимальним планом. Як показано на рисунку 5.1, цей план не змінюється при переході від c до c', і, навпаки, при переході від c до c'' лінія рівня цільової функції $L(x)=c''x$ перетинає лінію 2, що викличе зміну оптимального базисного плану, яким тепер стане точка x'. Використовуючи умови оптимальності плану ЗЛП

$$\Delta_j = L_j - c_j \geq 0, \tag{5.8}$$

можна одержати кількісні оцінки для меж коливань коефіцієнтів цільової функції, при яких не відбувається зміна оптимального плану. Припустимо, що варіації піддався певний елемент $c_r' = c_r + \Delta c_r$.

Можливі два випадки:

1. Стовпець r не входить до оптимального базису. Тоді для незмінності оптимального плану необхідно й достатньо виконання умови

$$\Delta_r' = L_r - c_r' \geq 0.$$

Звідси можна одержати значення для припустимої варіації

$$\Delta c_r \leq L_r - c_r. \qquad (5.9)$$

2. Стовпець r входить до оптимального базису. У цьому випадку для збереження оптимальності поточного плану потрібно виконання для всіх небазисних стовпців умов (5.8), або

$$\Delta_j = \sum_{i=1}^{m} c_{j_{баз}} a_{ij} - c_j \geq 0, \qquad (5.10)$$

оскільки у цьому випадку зміни відбуваються також у стовпчику $C_{баз}$ симплекс-таблиці, а це, у свою чергу, стосується всіх ненульових оцінок Δ_j.

Отже, у цьому випадку припустима варіація має задовольняти умовам

$$\Delta c_r \leq \sum_{i=1}^{m} c_{j_{баз}} a_{ij} - c_j. \qquad (5.11)$$

Повернемося до числового прикладу та визначимо межі зміни параметрів цільової функції, при яких знайдений план $x^*=(83;\ 36;\ 0;\ 0;\ 21)$ залишається оптимальним. У даній задачі інтерес представляють варіації коефіцієнтів c_1 і c_2, які стоять при базисних змінних в оптимальному плані.

Запишемо умови (5.11) для коефіцієнта c_1

$$\begin{cases} \Delta_3 = L_3 - c_3 = (6+\Delta c_1)*4/7 + 5*(-3/7) + 0*11/28 - 0 = \dfrac{24}{7} + \dfrac{4}{7}\Delta c_1 - \dfrac{15}{7} \\ \Delta_4 = L_4 - c_4 = (6+\Delta c_1)*(-3/7) + 5*4/7 + 0*11/7 - 0 = -\dfrac{18}{7} - \dfrac{3}{7}\Delta c_1 + \dfrac{20}{7} \end{cases} \Rightarrow$$

$$\begin{cases} \Delta_3 = \dfrac{24}{7} + \dfrac{4}{7}\Delta c_1 - \dfrac{15}{7} \geq 0 \\ \Delta_4 = -\dfrac{18}{7} - \dfrac{3}{7}\Delta c_1 + \dfrac{20}{7} \geq 0 \end{cases} \Rightarrow \begin{cases} \Delta c_1 \geq -\dfrac{9}{4} \\ \Delta c_1 \leq \dfrac{2}{3} \end{cases},$$

$$-\frac{9}{4} \leq \Delta c_1 \leq \frac{2}{3},$$

$$3{,}75 \leq c_1 \leq 6{,}67.$$

Аналогічно визначимо варіацію коефіцієнта c_2.

$$\begin{cases} \Delta_3 = L_3 - c_3 = 6*4/7 + (5+\Delta c_2)*(-3/7) + 0*11/28 - 0 = \dfrac{9}{7} - \dfrac{3}{7}\Delta c_2 \\ \Delta_4 = L_4 - c_4 = 6*(-3/7) + (5+\Delta c_2)*4/7 + 0*(-11/7) - 0 = \dfrac{2}{7} + \dfrac{4}{7}\Delta c_2 \end{cases} \Rightarrow$$

$$\begin{cases} \Delta_3 = \dfrac{9}{7} - \dfrac{3}{7}\Delta c_2 \geq 0 \\ \Delta_4 = -\dfrac{2}{7} + \dfrac{4}{7}\Delta c_2 \geq 0 \end{cases} \Rightarrow \begin{cases} \Delta c_2 \leq 3 \\ \Delta c_2 \geq \dfrac{1}{2} \end{cases},$$

$$\frac{1}{2} \le \Delta c_2 \le 3,$$

$$5{,}5 \le c_2 \le 8.$$

Наведений приклад дослідження чутливості оптимального плану стосовно зміни параметрів задачі є простим. Існують і складніші задачі, у яких, наприклад, досліджуються спільні варіації параметрів різних типів. Вони складають предмет спеціального розділу дослідження операцій, що одержав назву параметричного програмування.

5.3 Оцінка рентабельності виробленої продукції

Оцінка рентабельності продукції, що випускається підприємством, базується на теоремі рівноваги, (теоремі 4.4) і виконується за допомогою двоїстих оцінок і обмежень двоїстої задачі, що характеризують кожний вид продукції. З цієї теореми випливає, що кількість продукції, витрати на виробництво якої перевищують дохід, в оптимальному плані дорівнює нулю.

Якщо при підстановці u^* до системи обмежень двоїстої задачі вартість ресурсів, витрачених на одиницю продукції (ліва частина), перевищує ціну цієї продукції (права частина), то виробництво такої продукції для підприємства недоцільно. Тобто у цьому випадку даний вид продукції є нерентабельним. Якщо ж співвідношення виконується як рівність, продукція є рентабельною.

Аналогічні результати можна одержати, проаналізувавши симплекс-різниці Δ_j у стовпчиках, що відповідають досліджуваним видам продукції. Їхні значення показують, наскільки вартість ресурсів перевищує ціну одиниці відповідної продукції. Так, якщо симплекс-різниця дорівнює нулю $\Delta_j = 0$, то продукція рентабельна. І, навпаки, якщо $\Delta_j \ne 0$, то відповідна продукція нерентабельна. Помітимо також, що індексний рядок останньої симплекс-таблиці містить значення додаткових змінних двоїстої задачі u^*. Таким чином, якщо вони перевищують ціну продукції відповідного виду, то цей вид продукції є рентабельним.

Звернемося до числового прикладу. Підставимо отримані тіньові оцінки в обмеження двоїстої задачі:

$$\begin{cases} 4*\dfrac{9}{7}+3*\dfrac{2}{7}+3*0=6 \quad - \quad \textit{продукція A є рентабельною,} \\ 3*\dfrac{9}{7}+4*\dfrac{2}{7}+5*0=5 \quad - \quad \textit{продукція B є рентабельною.} \end{cases}$$

Оскільки обоє обмеження виконуються як строгі рівності, обидва види продукції А і В є рентабельними. Це підтверджується й тим, що в оптимальному плані $x^*=(83;\ 36;\ 0;\ 0;\ 21)$ обидва елементи x_1 і x_2, що відповідають обсягам виробництва, не нульові.

Проаналізуємо додаткові змінні двоїстої задачі, які розміщуються в індексному рядку симплекс-таблиці в стовпчиках A_1 і A_2. Їхні оптимальні значення $u_4=6$; $u_5 = 5$. Відповідно симплекс-різниці $\Delta_1 = \Delta_2 = 0$, що також свідчить про рентабельність продукції А і В.

5.4 Аналіз обмежень дефіцитних і недефіцитних ресурсів

Статус ресурсів прямої задачі можна визначити трьома способами. Перший - підстановкою x^* до системи обмежень прямої задачі. Якщо обмеження виконується як рівність, то відповідний ресурс є дефіцитним, у протилежному випадку - недефіцитним.

Другий спосіб - за допомогою додаткових змінних прямої задачі. Якщо додаткова змінна в оптимальному плані дорівнює нулю, то відповідний ресурс дефіцитний, а якщо відмінна від нуля - ресурс недефіцитний.

Третій спосіб - за допомогою двоїстих оцінок. Якщо $u_i \neq 0$, то зміна (збільшення або зменшення) обсягів i-го ресурсу призводить до відповідної зміни доходу підприємства, і тому такий ресурс є дефіцитним. Якщо $u_i=0$, то i-й ресурс недефіцитний.

Звернемося до числового прикладу. Підставимо компоненти оптимального плану $x^*=(83; 36; 0; 0; 21)$ до системи обмежень прямої задачі:

$$\begin{cases} 4*83+3*36=440 & \textit{ресурс витрачено повністю,} \\ 3*83+4*36=393 & \textit{ресурс витрачено повністю,} \\ 3*83+5*36=429 & \textit{ресурс витрачено не повністю.} \end{cases}$$

Змінні u_1 = 1,286 і u_2 = 0,286 є умовними двоїстими оцінками одиниці сировини S_1 і S_2 відповідно. Ці оцінки відмінні від нуля, а сировина S_1 і S_2 повністю використовується при оптимальному плані виробництва виробів А і В. Двоїста оцінка одиниці сировини S_3 u_3=0. Цей вид сировини не використовується повністю при оптимальному плані виробництва.

Таким чином, додатну двоїсту оцінку мають лише ті види сировини, які повністю використовуються при оптимальному плані виробництва. Тому двоїсті оцінки визначають дефіцитність використовуваної підприємством сировини.

$$\begin{cases} u_1=1{,}286 & \textit{ресурс } 1 \textit{ дефіцитний,} \\ u_2=0{,}286 & \textit{ресурс } 2 \textit{ дефіцитний,} \\ u_3=0 & \textit{ресурс } 3 \textit{ недефіцитний.} \end{cases}$$

Контрольні запитання

1. У чому полягає економічна інтерпретація прямої та двоїстої задач лінійного програмування?
2. Як визначити, є ресурс дефіцитним чи ні?
3. Як визначити, що продукція є рентабельною або нерентабельною?
4. У чому полягає економічний зміст змінних двоїстої задачі?
5. Який зміст вкладають у поняття «параметрична стійкість»?
6. Сформулюйте умови для припустимих змін цільової функції задачі, при яких її оптимальний план залишається незмінним.
7. Як визначити статус ресурсів прямої задачі?
8. Як визначити інтервали стійкості двоїстих оцінок щодо зміни запасів дефіцитних ресурсів?
9. Як визначити оптимальний план виробництва продукції та зміну доходу підприємства при збільшенні або зменшенні обсягу ресурсів?
10. Як розрахувати інтервали можливої зміни ціни одиниці кожного виду продукції?

ЗМ 3 Методи розв’язання нелінійних оптимізаційних задач

ТЕМА 6
ЦІЛОЧИСЛОВІ ТА ДРІБНО-ЛІНІЙНІ ЗАДАЧІ ЛІНІЙНОГО ПРОГРАМУВАННЯ

6.1 Типи задач дискретного програмування

Цілочислове лінійне програмування орієнтоване на розв’язання задач лінійного програмування, у яких всі або деякі змінні повинні приймати цілочислові (або дискретні) значення. Багато економічних задач характеризуються тим, що обсяги керованих ресурсів можуть приймати тільки цілі значення. До цілочислового програмування належать так само задачі, у яких змінні можуть приймати тільки два значення – 0 або 1 (булеві або бінарні змінні). До задач цілочислового програмування належать задачі про призначення, про найкоротший шлях та інші. У загальному вигляді задачу цілочислового програмування можна сформулювати як задачу знаходження максимуму (або мінімуму) цільової функції $L(x_1, x_2,..., x_n)$ на множині D, зумовленій системою обмежень. Загальна задача цілочислового програмування формулюється в такий спосіб

Знайти найбільше або найменше значення лінійної функції

$$L = \sum_{j=1}^{n} c_j x_j \to extr \tag{6.1}$$

за умови

$$\sum_{j=1}^{n} a_{ij} x_j = b_i, \quad i = \overline{1,m}, \tag{6.2}$$

$$x_j \geq 0, \quad x_j - цілі, \quad j = \overline{1,n}..$$

Умова $x_j - цілі$ називається умовою дискретності. Особливе місце серед дискретних задач займає цілочислова задача лінійного програмування в канонічній формі (ЦКЗЛП). У деяких ситуаціях вимогу «цілочислове» можна накласти лише на деякі змінні x_j, що кардинально не змінює характеру задачі.

Принципова складність, викликувана наявністю умов цілочисленості в системі обмежень оптимізаційної задачі, полягає в тому, що в значній кількості випадків неможливо замінити дискретну задачу її безперервним аналогом і, знайшовши відповідний розв’язок, округлити його компоненти до найближчих цілих значень. На рисунку 6.1 показано, що при округленні оптимального плану x^* звичайної задачі ЛП до цілих значень виходить точка B, що не належить області припустимих планів. Якщо навіть оптимальний план безперервної задачі, округлений до цілих значень компонент, виявиться припустимим, то цільова функція може повести себе так, що її значення буде на ньому істотно «гірше», ніж на оптимальному плані цілочислової задачі.

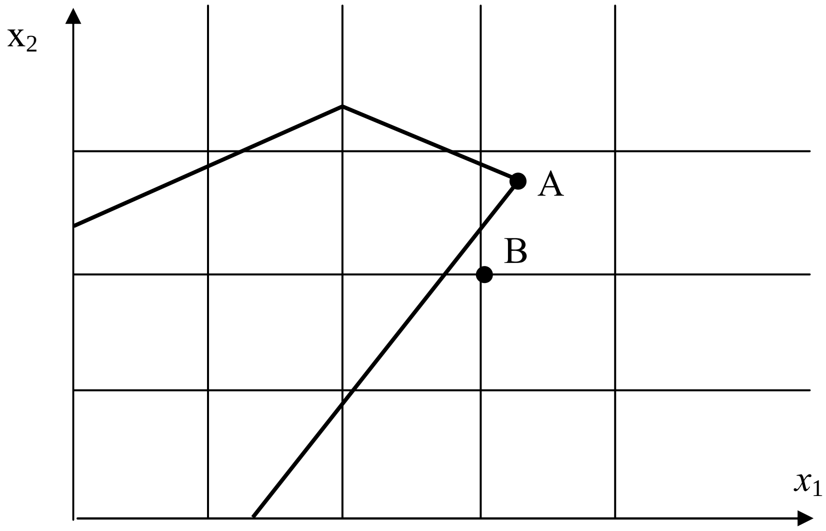

Рисунок 6.1 – Розв’язання цілочислової задачі

Перелічені проблеми визначили необхідність розробки спеціальних методів розв’язання дискретних і цілочислових задач.

У літературі, як правило, виділяють наступні класи дискретних оптимізаційних задач:

- задача з неділимостями;
- екстремальні комбінаторні задачі;
- задачі з розривними цільовими функціями;
- задачі на незв'язних і неопуклих областях та ін.

Задача з неділимостями. У переважній більшості випадків наявність умов неділимості визначається фізичними властивостями об'єктів, наприклад, вони можуть з'явитися як додаткові обмеження в задачі виробничого планування, якщо в ній здійснюється управління випуском дуже крупної продукції. Класичним представником задач даного класу є так звана задача про ранець. Вона полягає в тому, що солдат (або турист), що збирається в похід, може нести вантаж вагою, що не перевищує *W* кг. Цей вантаж може складатися з набору предметів n типів, кожний предмет j-го типу важить *w* кг і характеризується певною «корисністю». Треба визначити, скільки предметів кожного виду потрібно покласти в ранець, щоб його сумарна корисність була максимальною. До такого формулювання можуть бути зведені багато економічних задач. У літературі ця задача також відома як задача про завантаження судна.

Комбінаторні задачі. До даного класу належать задачі оптимізації функції, заданої на кінцевій множині, елементами якої є вибірки з n об'єктів. Класичним представником такої задачі є задача про комівояжера. Вона полягає в складанні маршруту відвідування торговельним агентом, що знаходиться в певному початковому пункті, n інших міст за умови, що задано матрицю вартостей переїздів з міста до міста. Причому припустимим є маршрут, що передбачає однократне відвідування всіх міст і повернення у вихідний пункт.

Найкращий маршрут має мінімізувати сумарну вартість переїздів. Кожний припустимий маршрут можна ототожнити з перестановкою n чисел. Задача комівояжера має велику кількість змістовних аналогів. Зокрема, задача розробки графіка переналагодження обладнання, що може випускати різні типи виробів, але вимагає певних витрат (часових або матеріальних) при переході з одного технологічного режиму на іншій.

Задача з розривними цільовими функціями. Багато економічних систем характеризуються наявністю так званих постійних витрат, які повинні бути зроблені незалежно від обсягу виробництва. Урахування у моделях цих і подібних факторів призводить до появи в них цільових функцій, що не є безперервними. Як приклад можна навести транспортну задачу з фіксованими доплатами. У цьому випадку цільова функція сумарних витрат на перевезення містить «стрибкоподібні» розриви, що істотно утруднює її мінімізацію.

Методи розв'язання задач цілочислового лінійного програмування базуються на використанні обчислювальних можливостей методів лінійного програмування. Зазвичай алгоритми цілочислового програмування включають три кроки:

1. «Ослаблення» простору припустимих розв'язків задачі шляхом відкидання вимоги цілочисленості. У результаті виходить звичайна задача лінійного програмування.
2. Визначення оптимального розв'язку задачі лінійного програмування.
3. Маючи отриманий оптимальний розв'язок, додають спеціальні обмеження, які ітераційним шляхом змінюють простір припустимих розв'язків задачі лінійного програмування таким чином, щоб в остаточному підсумку вийшов оптимальний розв'язок, що задовольняє вимогам цілочисленості.

Розроблено два методи побудови спеціальних обмежень:

- метод Гоморі (або метод площин, що відтинають, він був уперше запропонованим Р. Гоморі у 1957-1958 р.);
- метод віток і границь.

Помітимо також, що досить ефективний і широко застосовуваний підхід до розв'язання цілочислових задач ґрунтується на зведенні їх до задач транспортного типу. Це пояснюється тим, що якщо в умовах транспортної задачі значення запасів і потреб є цілочисловими, то цілочисловим буде й оптимальний план.

6.2 Метод Гоморі

Сутність методу Гоморі полягає в тому, що спочатку задача вирішується без урахування умови цілочисленості. Якщо отриманий у результаті оптимальний план x^* містить тільки цілі компоненти, задачу вирішено. У протилежному випадку до системи обмежень задачі додають нове обмеження, що має наступні властивості:

- воно має бути лінійним;
- повинне відтинати знайдений нецілочисленний оптимальний план x^*;
- не повинне відтинати жодного цілочислового плану.

Додаткове обмеження, що має зазначені властивості, називається **правильним відсіканням**. Далі задача вирішується з урахуванням нового обмеження. Після цього, якщо буде потреба, додається ще одно обмеження і т. д.

Геометрично додавання кожного лінійного обмеження відповідає проведенню прямої (гіперплощини), що відтинає від багатокутника (багатогранника) розв’язків певну його частину разом з оптимальною точкою з нецілими координатами, але не зачіпає ні однієї з цілих точок цього багатогранника. У результаті новий багатогранник розв’язків містить всі цілі точки, що містились в первісному багатограннику розв’язків, і відповідно отриманий при цьому багатограннику оптимальний розв’язок буде цілочисловим.

Один з алгоритмів розв’язання задачі лінійного цілочислового програмування, запропонований Гоморі, ґрунтується на симплексному методі і використовує досить простий спосіб побудови правильного відсікання.

Нехай задача лінійного програмування має кінцевий оптимум і на останньому кроці її розв’язання симплексним методом отримані наступні рівняння, у яких базисні змінні x_1, x_2, …, x_m виражені через вільні змінні x_{m+1}, x_{m+2}, …, x_n оптимального розв’язку

$$\begin{aligned} x_1 &= b_1 - a_{1\,m+1}x_{m+1} - \ldots - a_{1n}x_n, \\ x_2 &= b_2 - a_{2\,m+1}x_{m+1} - \ldots - a_{2n}x_n, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_m &= b_m - a_{m\,m+1}x_{m+1} - \ldots - a_{mn}x_n, \end{aligned} \tag{6.3}$$

де оптимальним розв’язком задачі є $x^* = (b_1, b_2, \ldots, b_m, 0, 0, \ldots, 0)$, у якому, наприклад, b_i – нецілий компонент.

Нагадаємо, що цілою частиною числа b називається найбільше ціле число $[b]$, що не перевершує b, а дробовою частиною числа b - число $\{b\}$, що дорівнює різниці між цим числом і його цілою частиною, тобто $\{b\} = b - [b]$.

Сформувавши за i-м рівнянням системи (6.3) додаткове обмеження

$$\{b_i\} - \{a_{i\,m+1}\}x_{m+1} - \ldots - \{a_{in}\}x_n \le 0, \tag{6.4}$$

можна довести, що воно має всі властивості правильного відсікання.

Для розв’язання задачі цілочислового програмування методом Гоморі використовується наступний алгоритм:

1. Вирішити задачу симплексним методом без урахування умови цілочисленості. Якщо всі компоненти оптимального плану цілі, то він є оптимальним і для задачі цілочислового програмування. Якщо перша задача є нерозв'язною (тобто не має кінцевого оптимуму або умови її суперечливі), то й друга задача так само нерозв'язна.

2. Якщо серед компонентів оптимального розв’язку є нецілі, то вибрати компоненту з найбільшою цілою частиною і за відповідним їй рівнянням системи (6.3) сформувати правильне відсікання (6.4).

3. Отриману нерівність перетворити на рівносильне рівняння шляхом введення додаткової невід'ємної цілочислової змінної

$$\{b_i\}-\{a_{i\,m+1}\}x_{m+1}-\ldots-\{a_{in}\}x_n+x_{n+1} \qquad (6.5)$$

і включити його до системи обмежень. Необхідно пам'ятати, що після включення до системи обмежень додаткового рівняння, яке відповідає правильному відсіканню, завжди буде виходити неприпустимий базисний розв'язок. Для одержання припустимого базисного розв'язку потрібно перевести в базисні змінні одну з вільних змінних.

4. Отриману розширену задачу вирішити симплексним методом. Якщо знайдений оптимальний план буде цілочисловим, то задача цілочислового програмування вирішена. У протилежному випадку треба повернутися до п.2 алгоритму.

Якщо задачу розв'язано в цілих числах, то після кінцевого числа кроків оптимальний цілочисловий план буде знайдений.

Якщо в процесі розв'язання з'явиться рівняння (що виражає базисну змінну через вільні) з нецілим вільним членом і цілими іншими коефіцієнтами, то відповідне рівняння не має розв'язання в цілих числах. У цьому випадку і задача не має цілочислового оптимального розв'язку.

Розглянемо особливості застосування методу Гоморі на прикладі. Нехай дано задачу з наступними умовами:

$$L=3x_1+x_2+2x_3 \to max,$$
$$x_1+2x_2+4x_3=9,$$
$$3x_1+x_2+5x_3=10,$$
$$x_j \ge 0,\quad x_j - цілі,\quad j=\overline{1,3}.$$

Використовуючи звичайний симплекс-алгоритм, вирішуємо безперервний аналог вихідної задачі, у якої ігноруються умови цілочисловості. Як вихідний базис можна взяти перший і другий стовпці. На його основі заповнюється симплекс-таблиця 6.1.

Таблиця 6.1 – Вихідна симплекс-таблиця

Базис	$C_{баз}$	C_j	3	1	2
		B	A_1	A_2	A_3
A_1	3	11/5	1	0	6/5
A_2	1	17/5	0	1	7/5
		L_j: 10	3	1	5
		Δ_j	0	0	3

Як видно з рядка оцінок, даний базис є оптимальним, проте відповідний йому план $x^*=\left(\frac{11}{5};\frac{17}{5};0\right)$ не є цілочисловим, тому обираємо рядок, що містить перший нецілий елемент, і відповідно до формули (6.4) будуємо обмеження, що відтинає:

$$\left\{\frac{11}{5}\right\}-\left\{\frac{6}{5}\right\}x_3 \le 0.$$

Увівши додаткову цілочислову змінну $x_4 \geq 0$, одержимо рівносильне нерівності рівняння

$$-\frac{1}{5}x_3 + x_4 = -\frac{1}{5},$$

яке необхідно включити до системи обмежень. Уведемо його до симплекс-таблиці 6.2.

Таблиця 6.2 - Введення рівняння до симплекс-таблиці

Базис	$C_{баз}$	C_j	3	1	2	0
		B	A_1	A_2	A_3	A_4
A_1	3	11/5	1	0	6/5	0
A_2	1	17/5	0	1	7/5	0
A_4	0	-1/5	0	0	-1/5	1
	L_j	10	3	1	5	0
	Δ_j		0	0	3	0

Врахуємо, що отриманий базисний розв'язок є неприпустимим (компонента b_3 від'ємна). Для одержання припустимого базисного розв'язку переведемо в базисні змінні вільну змінну x_3 (табл. 6.3).

Таблиця 6.3 - Введення вільної змінної x_3

Базис	$C_{баз}$	C_j	3	1	2	0
		B	A_1	A_2	A_3	A_4
A_1	3	1	1	0	0	6
A_2	1	2	0	1	0	7
A_3	2	1	0	0	1	-5
	L_j	7	3	1	2	15
	Δ_j		0	0	0	15

Отриманий план $x^*=(1, 2, 1, 0)$ є не тільки оптимальним (всі $b_j>0$), але й цілком складається з цілочислових компонент, тобто розв'язок задачі знайдено, max L = 7.

6.3 Метод віток і границь

Метод віток і границь - один з комбінаторних методів. Його сутність полягає в упорядкованому переборі варіантів і розгляданні тих з них, які за певними ознаками є перспективними. Уперше даний метод запропонували в 1960 р. Ленг і Дойг, а у 1963 р. він отримав розвиток у зв'язку з виходом роботи Літтла, Мурті, Суїні та Керел, присвяченої розв'язанню задачі про комівояжера.

Метод віток і границь полягає в тому, що множина припустимих розв'язків (планів) певним способом розбивається на підмножини, кожна з яких цим самим способом знову розбивається на підмножини. Процес триває доти, поки не буде отриманий оптимальний цілочисловий розв'язок вихідної задачі.

Якщо задача максимізації лінійної функції L вирішена симплексним методом без урахування цілочисленості змінних, то стають відомими нижня та верхня границі для кожної цілочислової змінної x_j: $[x_j]<x_j<[x_j]+1$ і нижня границя лінійної функції L_0, тобто при будь-якому плані x $L(x) \geq L_0$. Тоді з області припустимих значень, наприклад, змінної x_r виключається область $[x_r] < x_r < [x_r]+1$, де $[x_r]$ – ціла частина числа x_r. У результаті це дозволяє сформувати дві задачі, що відрізняються тим, що в одну з них додано обмеження $x_r \leq [x_r]$, а в іншу – обмеження $x_r \geq [x_r]+1$.

Сформовані задачі вирішують у будь-якому порядку. Залежно від отриманого розв'язку список таких задач може розширюватися або зменшуватися. Якщо в ході розв'язання будь-якої з задач отриманий нецілочисловий оптимальний план, для якого $L(x) \leq L_0$, то дану задачу виключають із списку. Якщо $L(x) \geq L_0$, то з даної задачі формують дві нові задачі.

Розглянемо приклад:

$$L = 5x_1 + 4x_2 \rightarrow max,$$

$$x_1 + x_2 \leq 5,$$

$$10x_1 + 6x_2 \leq 45,$$

$$x_j \geq 0, \quad x_j - цілі, \quad j = \overline{1,2}.$$

Вирішимо задачу шляхом відкидання умов цілочисленості. Скористаємось функцією Excel «Пошук рішення» і одержимо

$$x_1=3,75,\ x_2=1,25,\ L=23,75.$$

Оберемо одну з цілочислових змінних, значення якої в оптимальному розв'язку не є цілочисловим, наприклад, x_1. Очевидно, що область $3<x_1<4$ простору припустимих розв'язків не містить цілочислових значень змінної x_1, і отже може бути виключена з розгляду. Це еквівалентно заміні вихідної задачі двома новими задачами. В одну з них додамо обмеження $x_1 \leq 3$, одержимо

$$L = 5x_1 + 4x_2 \rightarrow max,$$

$$x_1 + x_2 \leq 5,$$

$$10x_1 + 6x_2 \leq 45,$$

$$x_1 \leq 3,$$

$$x_j \geq 0, \quad x_j - цілі, \quad j = \overline{1,2}.$$

Використовуючи функцію Excel «Пошук рішення», отримаємо оптимальний цілочисловий план

$$x_1=3;\ x_2=2;\ L=23.$$

Якість отриманого цілочислового розв'язку оцінити неможливо, тому що розв'язок другої задачі може призвести до більшого значення цільової функції. Це треба перевірити. Сформуємо другу задачу, додавши до неї обмеження $x_1 \geq 4$:

$$L = 5x_1 + 4x_2 \rightarrow max,$$

$$x_1 + x_2 \leq 5,$$

$$10x_1 + 6x_2 \leq 45,$$

$$x_1 \geq 4,$$

$$x_j \geq 0, \quad x_j - цілі, \quad j = \overline{1,2}.$$

В оптимальному розв’язку другої задачі змінна x_2 не є цілим числом:

$$x_1=4,\ x_2=0{,}83,\ L=23{,}33.$$

Оскільки значення змінної $x_2=0{,}83$ не є цілим числом, другу задачу необхідно досліджувати далі, розділивши її у свою чергу ще на дві задачі і т. д. Проілюструємо хід розв’язання схемою, наведеною на рисунку 6.2.

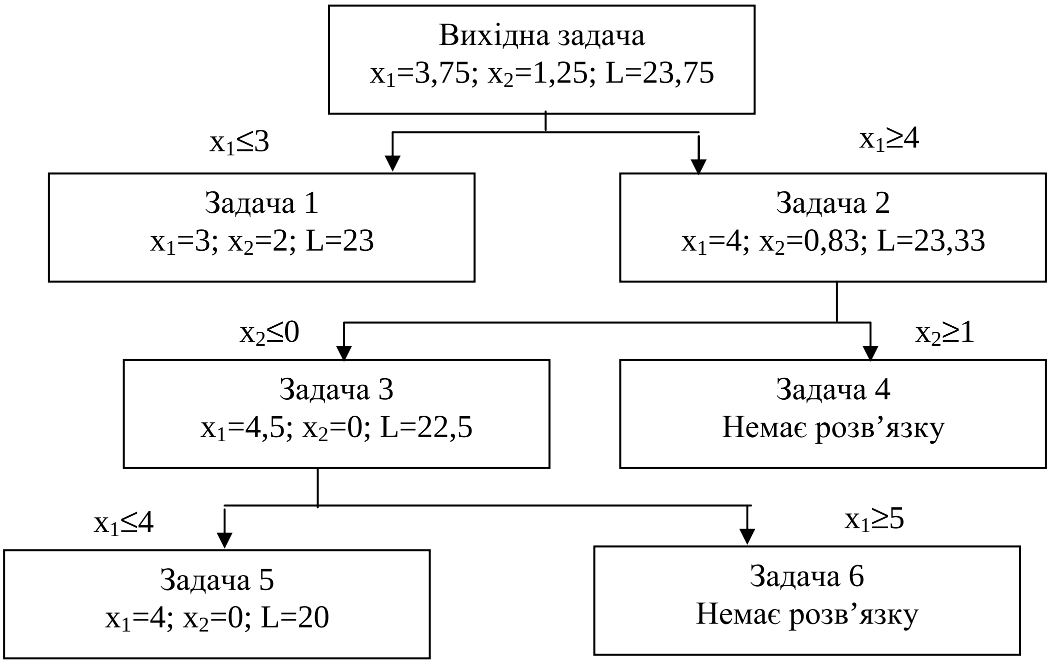

Рисунок 6.2 – Схема методу віток і границь

У ході розв’язання можна було спочатку як змінну розгалуження прийняти x_2. При цьому наступні обчислення можуть істотно відрізнятися.

Очевидним недоліком алгоритму методу віток і границь при розв’язанні задач великої розмірності є необхідність перебрати занадто велику кількість варіантів, перш ніж буде знайдений оптимальний план.

6.4 Дрібно-лінійне програмування

В економічних задачах оптимізації цільову функцію можна представити у вигляді дрібно-лінійної функції. Це має місце, зокрема, якщо критерієм оптимальності є продуктивність праці або показник рентабельності та ін. У цих випадках загальна оптимізаційна модель має вигляд:

Знайти
$$Z=\frac{\sum_{j=1}^{n}c_jx_j+c_0}{\sum_{j=1}^{n}d_jx_j+d_0}\to extr \tag{6.5}$$

за умови

$$\begin{aligned}&\sum_{j=1}^{n}a_{ij}x_j=b_i,\quad i=\overline{1,m},\\ &x_j\ge 0,\quad j=\overline{1,n},\end{aligned} \tag{6.6}$$

причому, знаменник дробу не може дорівнювати нулю в області припустимих розв'язків системи обмежень.

Для розв'язання задачі дрібно-лінійного програмування її зводять до задачі лінійного програмування шляхом заміни змінних. Позначимо

$$\sum_{j=1}^{n} d_j x_j + d_0 = \frac{1}{s_0}, \tag{6.7}$$

тоді $s_j = s_0 x_j$, $j = \overline{1,n}$. Одержимо перетворену модель задачі лінійного програмування:

Знайти
$$L = \sum_{j=1}^{n} c_j s_j + x_0 s_0 \to extr \tag{6.8}$$

за умови

$$\sum_{j=1}^{n} a_{ij} s_j - b_i s_0 = 0, \quad i = \overline{1,m},$$
$$\sum_{j=1}^{n} d_j s_j + d_0 s_0 = 1, \tag{6.9}$$
$$s_j \ge 0, \quad j = \overline{1,n}, \quad s_0 > 0.$$

Після визначення оптимального плану задачі лінійного програмування $s^*=(s_1, s_2, \ldots, s_n)$, наприклад, за допомогою симплекс-методу, компоненти оптимального плану дрібно-лінійної задачі визначають за формулою

$$x_j = \frac{s_j}{s_0}, \quad j = \overline{1,n}. \tag{6.10}$$

Контрольні запитання

1. Які основні проблеми виникають під час розв'язання дискретних задач?
2. Сформулюйте задачу про ранець.
3. Які економіко-математичні моделі можна звести до задачі про комівояжера?
4. Наведіть приклад моделей з розривними цільовими функціями.
5. Який принцип використовується для побудови правильного відсікання в методі Гоморі?
6. Яку роль відіграє алгоритм двоїстого симплекс-методу при розв'язанні цілочислової лінійної задачі за методом Гоморі?
7. Перелічіть принципові ідеї, що лежать в основі методів віток і границь.
8. Як провадиться побудова відсікання при розв'язанні цілочислової лінійної задачі за методом віток і границь?
9. Опишіть схему розв'язання цілочислової задачі лінійного програмування за методом віток і границь.
10. За рахунок яких перетворень вдається побудувати сполучений базис при додаванні обмеження, що відтинає?
11. Сформулюйте задачу дрібно-лінійного програмування.
12. Охарактеризуйте метод розв'язання дрібно-лінійних задач.

ТЕМА 7
НЕЛІНІЙНІ ОПТИМІЗАЦІЙНІ МОДЕЛІ

7.1 Постановка задачі нелінійного програмування (ЗНП)

Залежності між керованими змінними далеко не завжди можна описати за допомогою адекватної лінійної моделі. Наприклад, у лінійних моделях ціна товару вважається незалежною від кількості реалізованого продукту, у той час як вона може залежати від обсягу партії товару. З приводу технологічних обмежень можна помітити, що витрата певних видів сировини та ресурсів відбувається не лінійно, а стрибкоподібно (залежно від обсягу виробництва). Спроби врахувати ці фактори призводять до формулювання більш загальних і складних оптимізаційних задач. Вивчення методів їх розв’язання складає предмет наукової області, що одержала назву нелінійного програмування.

Для задач нелінійного програмування не існує універсального методу розв’язання, тому щоразу необхідно доводити існування розв’язку задачі, а також його одиничність. Відомі точні методи розв’язання нелінійних задач, але алгоритми їх розв’язання є трудомісткими навіть для сучасного програмного забезпечення ЕОМ. На практиці частіше користуються наближеними методами, проблема яких пов'язана з пошуком локальних і глобальних оптимумів. Більшість наближених методів дозволяють визначити локальний оптимум. Визначивши всі локальні оптимуми та порівнявши їх, можна знайти глобальний оптимум. Але такий підхід не є ефективним для практичних розрахунків. Потрібно відзначити, що якщо в задачах лінійного програмування оптимальний розв’язок завжди знаходився на границі області обмежень, то в задачі нелінійного програмування він може знаходитись також і усередині цієї області.

У класичній теорії оптимізації для пошуку точок максимуму та мінімуму (екстремальних точок) функцій, як при відсутності, так і при наявності обмежень на змінні використовується апарат диференціального обчислення. Екстремальна точка функції $f(x)$ визначає або її максимальне, або мінімальне значення. З математичної точки зору точка $x_0=(x_1, x_2, \ldots, x_n)$ є точкою максимуму, якщо значення функції f в оточенні точки x_0 не перевищують $f(x_0)$. На рисунку 7.1 показані точки максимуму та мінімуму функції однієї змінної $f(x)$. Точки x_1, x_2, x_3, x_4 і x_6 складають множину екстремальних точок функції $f(x)$. Причому точка x_6 є точкою **глобального** (абсолютного) максимуму, тому що $f(x) = \max\{f(x_1), f(x_3), f(x_6)\}$, а точки $f(x_1)$ і $f(x_3)$ – точками **локального** (відносного) максимуму. Необхідною умовою існування екстремуму є рівність нулю похідних від $f(x)$. Але похідні дорівнюють нулю також і в точках перегину функції $f(x)$ у випадку однієї змінної (точка x_5), і в сідлових точках у випадку функції двох змінних. Тому рівність нулю похідних від $f(x)$ є необхідною, але недостатньою умовою існування екстремуму, а точки, у яких виконується дана умова, називаються **стаціонарними**.

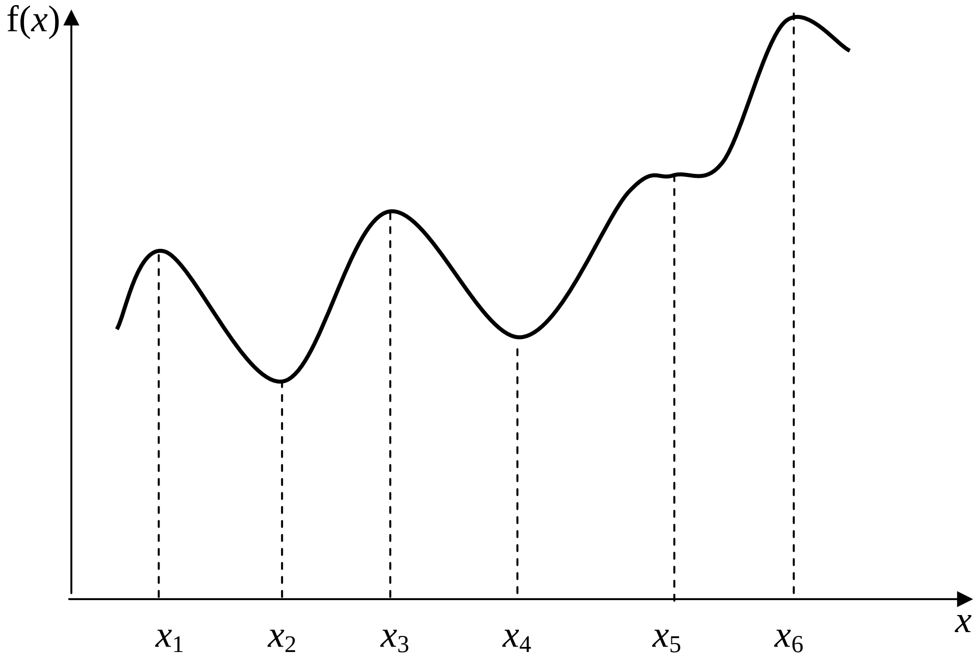

Рисунок 7.1 – Точки максимуму і мінімуму функції

Більшість методів, використовуваних для розв'язання задач нелінійного програмування, базуються на теорії диференціального обчислення. Серед них розрізняють прямі та непрямі методи. Прямими методами пошук оптимуму ведеться в напрямку найшвидшого зростання або убування цільової функції. До таких методів належать градієнтні методи, зокрема, метод найшвидшого спуску та метод припустимих напрямків. Непрямі методи передбачають перетворення вихідної задачі до вигляду, що дозволяє спростити пошук екстремуму цільової функції. До них належить квадратичне програмування та сепарабельне програмування, а також геометричне та стохастичне програмування.

Загальна задача нелінійного програмування (ЗЗНП) визначається як задача знаходження максимуму (або мінімуму) цільової функції $f(x_1, x_2,..., x_п)$ на множині D, зумовленій системою обмежень

$$\begin{gathered} g_1(x_1, x_2,..., x_n) \le 0, \\ g_2(x_1, x_2,..., x_n) \le 0, \\ \dots\dots\dots\dots \\ g_i(x_1, x_2,......., x_n) \le 0, \\ g_{i+1}(x_1, x_2,..., x_n) \le 0, \\ \dots\dots\dots\dots \\ g_m(x_1, x_2,......., x_n) = 0, \\ x_j \ge 0, \quad j = \overline{1, n}, \end{gathered} \tag{7.1}$$

де хоча б одна з функцій f або g_i є нелінійною.

Очевидно, що питання про тип оптимізації не є принциповим. Тому, для визначеності, надалі розглядатимемо задачі максимізації.

Як і в ЗЛП, вектор $x^*=(x_1, x_2,..., x_n)\in D$ називається припустимим планом, а якщо для будь-якого $x\in D$ виконується нерівність $f(x^*)>f(x)$, то x^* називають оптимальним планом. У цьому випадку x^* є точкою глобального максимуму.

З погляду економічної інтерпретації $f(x)$ може розглядатися як дохід, що одержує підприємство при плані випуску x, а $g_i(x)<0$ - як технологічні обмеження на можливості випуску продукції.

Набір обмежень, що визначають множину D, при необхідності завжди можна звести або до системи, що складається з одних нерівностей, або, додавши фіктивні змінні, до системи рівнянь. Перелічимо властивості ЗНП, які істотно ускладнюють процес їхнього розв’язання порівняно з задачами лінійного програмування.

1. Множина припустимих планів D може мати дуже складну структуру (наприклад, бути неопуклою або незв'язною).

2. Глобальний максимум (мінімум) може досягатися як усередині множини D, так і на її границях (де він не збігатиметься з жодним з локальних екстремумів).

3. Цільова функція f може бути недиференційованою, що утруднює застосування класичних методів математичного аналізу.

У чинність названих факторів задачі нелінійного програмування настільки різноманітні, що для них не існує загального методу розв’язання.

7.2 Класичний метод оптимізації з використанням множників Лагранжа

Одним з найбільш загальних підходів до розв’язання задач пошуку екстремуму функції при наявності сполучних обмежень на її змінні (задача умовної оптимізації) є метод Лагранжа. Багатьом студентам він має бути відомим з курсу диференціального обчислення. Ідея даного методу полягає у зведенні задачі пошуку умовного екстремуму цільової функції

$$f(x_1, x_2,..., x_n) \qquad (7.2)$$

на множині припустимих значень D, описуваній системою рівнянь

$$\begin{aligned} g_1(x_1, x_2,..., x_n)&= 0,\\ g_2(x_1, x_2,..., x_n)&= 0,\\ &\ldots\ldots\ldots\ldots\\ g_m(x_1, x_2,......., x_n)&= 0 \end{aligned} \qquad (7.3)$$

до задачі безумовної оптимізації функції

$$\Phi(x, \lambda) = f(x_1, x_2,..., x_n)+ \lambda_1 g_1(x_1, x_2,..., x_n) + ...+ \lambda_m g_m(x_1, x_2,..., x_n), \quad (7.4)$$

де λ_i - вектор додаткових змінних, що називаються невизначеними множниками Лагранжа.

Рівняння (7.3) називаються рівняннями зв'язку, а функція $\Phi(x, \lambda)$ - функцією Лагранжа.

Для функції $\Phi(x, \lambda)$ вирішують систему рівнянь

$$\begin{cases} \dfrac{\partial \Phi(x,\lambda)}{\partial x_j} = 0, (j = \overline{1,n}), \\ \dfrac{\partial \Phi(x,\lambda)}{\partial \lambda_i} = g_i(x) = 0, (i = \overline{1,m}) \end{cases} \tag{7.5}$$

щодо змінних x і λ.

Метод Лагранжа складається з наступних етапів.

1. Складання функції Лагранжа $\Phi(x, \lambda)$.

2. Знаходження частинних похідних

$$\frac{\partial \Phi(x,\lambda)}{\partial x_j},\ j = \overline{1,n} \quad \text{і} \quad \frac{\partial \Phi(x,\lambda)}{\partial \lambda_i},\ i = \overline{1,m}.$$

3. Розв’язання системи рівнянь (7.5) щодо змінних x і λ.

4. Дослідження точок, що задовольняють системі (7.5), на максимум (мінімум) за допомогою достатньої ознаки екстремуму.

Наявність останнього (четвертого) етапу пояснюється тим, що розглянутий алгоритм виконує необхідну, але не достатню умову екстремуму. Для визначення достатніх ознак умовного екстремуму і його типу існує спеціальний алгоритм, як правило, важко застосовний на практиці.

Основне практичне значення методу Лагранжа полягає в тому, що він дозволяє перейти від умовної оптимізації до безумовної. Проте задача розв’язання системи рівнянь (7.5), до якої збігається даний метод, у загальному випадку не простіше вихідної проблеми пошуку екстремуму (7.2)-(7.3). Методи, що припускають таке розв’язання, називаються непрямими. Вони можуть бути застосовані для досить вузького класу задач, для яких вдається одержати лінійну або таку, що збігається до лінійної, систему рівнянь (7.5). Їхнє застосування пояснюється необхідністю одержати розв’язок екстремальної задачі в аналітичній формі. При розв’язанні конкретних практичних задач зазвичай використовують прямі методи, засновані на ітеративних процесах обчислення та порівнянні значень функцій, що оптимізуються.

7.3 Опукле програмування

Основний недолік методів нелінійного програмування полягає в тому, що з їхньою допомогою не вдається знайти глобальний екстремум за наявності кількох локальних екстремумів. Теоретично нелінійне програмування розроблене тільки для одного окремого випадку опуклих функцій, і відповідно цей розділ названий опуклим програмуванням.

Функція $f(x_1, x_2,..., x_n)$ називається опуклою в області D, якщо для будь-яких двох точок $x^{(1)}, x^{(2)} \in D$ і будь-якого $\lambda = \overline{1,0}$ виконується нерівність

$$f((1-\lambda)x^{(1)} + \lambda x^{(2)}) \le (1-\lambda) f(x^{(1)}) + \lambda f(x^{(2)}), \tag{7.6}$$

якщо ж

$$f((1-\lambda)x^{(1)}+\lambda x^{(2)})\geq(1-\lambda)f(x^{(1)})+\lambda f(x^{(2)}),\tag{7.7}$$

то функція називається ввігнутою.

Геометричний зміст понять опуклості та увігнутості для функції однієї змінної представлений на рисунку 7.2. З нього видно, що графік опуклої функції лежить нижче відрізка, що з'єднує точки ($x^{(1)}$, $f(x^{(1)})$) і ($x^{(2)}$, $f(x^{(2)})$), а графік увігнутої - вище.

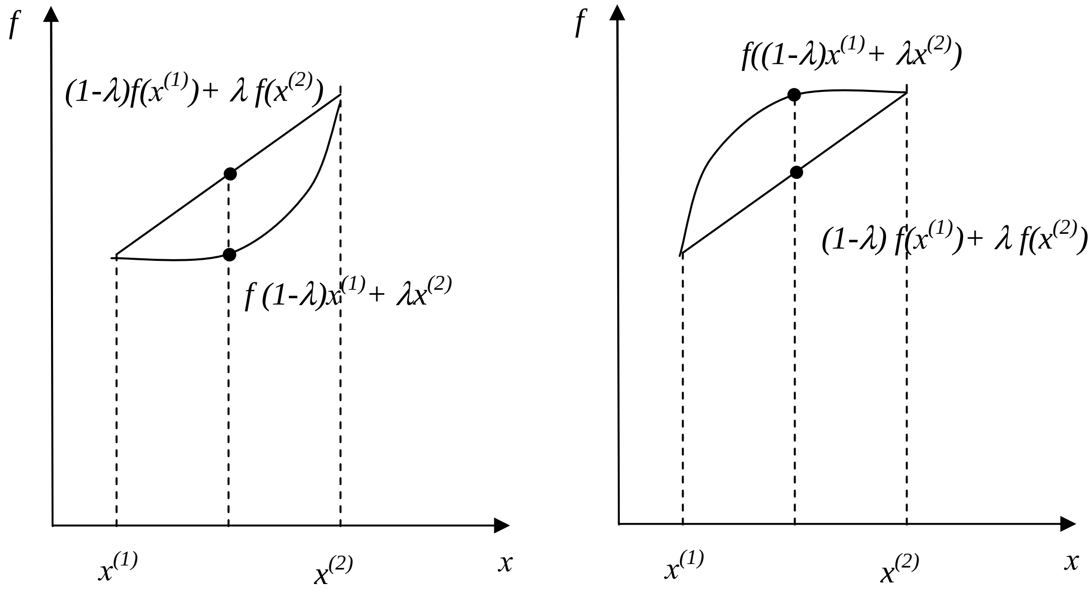

Рисунок 7.2 - Графіки опуклої та увігнутої функцій

Можна довести, що достатньою умовою опуклості функції $f(x_1, x_2,..., x_n)$ є додатна визначеність матриці

$$H(x)=\left(\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\right)_{n\times n},\tag{7.8}$$

що називається матрицею Гессе, у всіх точках $x\in D$. Відповідно, достатньою умовою ввігнутості є від'ємна визначеність матриці Гессе. Зокрема, для функцій однієї змінної достатньою умовою опуклості (увігнутості) є виконання нерівності $f^n(x)\geq 0$ ($f^n(x)\leq 0$).

Як видно з геометричної інтерпретації, для опуклої функції локальний екстремум, якщо він існує, збігається з глобальним. Для функції багатьох змінних точка x_0 являє собою вектор $f'(x)$ − вектор перших похідних (градієнт) функції $f(x)$, а матриця Гессе (гессіан) $f''(x)$ − симетричну матрицю других частинних похідних функції $f(x)$. Характер стаціонарної точки x_0 зв'язаний із знаковизначеністю матриці Гессе $f''(x^*)$.

Знаковизначеність матриці A залежить від знаків квадратичної форми

$$Q(x)=X^T AX=\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}x_i x_j\ .\tag{7.9}$$

Квадратична форма $Q(x)$ є додатно визначеною (напіввизначеною), якщо значення всіх кутових мінорів визначника $|A|$ додатні (невід'ємні). У цьому випадку матриця A називається додатно визначеною (напіввизначеною). k-м кутовим мінором визначника матриці $A_{n\times n}$ називається визначник виду

$$\begin{vmatrix} a_{11}\ a_{12} \dots a_{1k} \\ a_{21}\ a_{22} \dots a_{2k} \\ \vdots \quad \vdots \quad \dots \vdots \\ a_{k1}\ a_{k2} \dots a_{kk} \end{vmatrix}, \quad k = \overline{1,n}. \tag{7.10}$$

Квадратична форма $Q(x)$ є від'ємно визначеною, якщо значення k-х кутових мінорів визначника $|A|$ відмінні від нуля та мають знак $(-1)^k$. У цьому випадку матриця А називається від'ємно визначеною.

Квадратична форма $Q(x)$ є від'ємно напіввизначеною, якщо значення k-х кутових мінорів визначника $|A|$ дорівнюють нулю або мають знак $(-1)^k$.

Розглянемо приклад. Нехай є функція

$$f(x_1, x_2, x_3) = x_1 + 2x_3 + x_2 x_3 - x_1^2 - x_2^2 - x_3^2,$$

необхідно визначити її екстремум. Визначимо градієнт функції $f(x)$ (необхідна умова екстремуму)

$$\frac{\partial f}{\partial x_1} = 1 - 2x_1 = 0,$$

$$\frac{\partial f}{\partial x_2} = x_1 - 2x_2 = 0,$$

$$\frac{\partial f}{\partial x_3} = 2 + x_2 - 2x_3 = 0,$$

звідки одержимо стаціонарну точку $x_0 = \left(\frac{1}{2}, \frac{2}{3}, \frac{4}{3}\right)$. Визначимо характер стаціонарної точки. Для цього перевіримо достатність умови, склавши матрицю Гессе

$$H(x_0) = \begin{pmatrix} \frac{\partial^2 f}{\partial^2 x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_1 \partial x_3} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial^2 x_2} & \frac{\partial^2 f}{\partial x_2 \partial x_3} \\ \frac{\partial^2 f}{\partial x_3 \partial x_1} & \frac{\partial^2 f}{\partial x_3 \partial x_2} & \frac{\partial^2 f}{\partial^2 x_3} \end{pmatrix}_{x_0} = \begin{pmatrix} -2 & 0 & 0 \\ 0 & -2 & 1 \\ 0 & 1 & -2 \end{pmatrix}$$

і кутові мінори матриці $H(x_0)$

M_1= -2; M_2=(-2)*(-2)-0*0=4; M_3=-2*[(-2)*(-2)-1*1]=-6.

Отже, матриця $H(x_0)$ є від'ємно визначеною, звідки випливає, що функція $f(x)$ – увігнута, а отже стаціонарна точка $x_0 = \left(\frac{1}{2}, \frac{2}{3}, \frac{4}{3}\right)$ є точкою максимуму.

7.4 Необхідні й достатні умови існування сідлової точки. Теорема Куна-Таккера

Зупинимося на певних фундаментальних моментах теорії нелінійного програмування. Відправною точкою для них є поширення методу Лагранжа на розв'язання ЗНП з обмеженнями у формі нерівностей:

$$
\begin{gathered}
f(x_1, x_2, ..., x_n) \to max, \\
g_1(x_1, x_2, ..., x_n) \le 0, \\
g_2(x_1, x_2, ..., x_n) \le 0, \\
\ldots\ldots\ldots\ldots\ldots\ldots \\
g_m(x_1, x_2, ..., x_n) \le 0.
\end{gathered} \tag{7.11}
$$

Визначимо функцію Лагранжа:

$$\Phi(x,\lambda) = f(x_1, x_2, \ldots, x_n) - \sum_{i=1}^{m} \lambda_i g_i(x_1, x_2, \ldots, x_n). \tag{7.12}$$

Пара векторів (x, λ) називається сідловою точкою функції $\Phi(x, \lambda)$ у певній області, якщо для будь-яких x і λ, що належать цієї області,

$$\Phi(x, \lambda^*) \le \Phi(x^*, \lambda^*) \le \Phi(x^*, \lambda). \tag{7.13}$$

Нерівності (7.13) називають нерівностями сідлової точки.

Як приклад розглянемо функцію $\Phi(x,\lambda) = -x^2 + \lambda^2$. Сідловою точкою для цієї функції є точка (0, 0). Справді, $\Phi(0,0) = 0$, $\Phi(x,0) = -x^2$, $\Phi(0, \lambda) = \lambda^2$, а для будь-яких x і λ виконуються нерівності $-x^2 \le 0$ і $0 \le \lambda^2$.

На рисунку 7.3 зображений графік функції $\Phi(x,\lambda)$ (гіперболічний параболоїд), і, як видно, в оточенні точки (0,0) він дійсно за формою нагадує сідло, чим і пояснюється походження відповідного терміна.

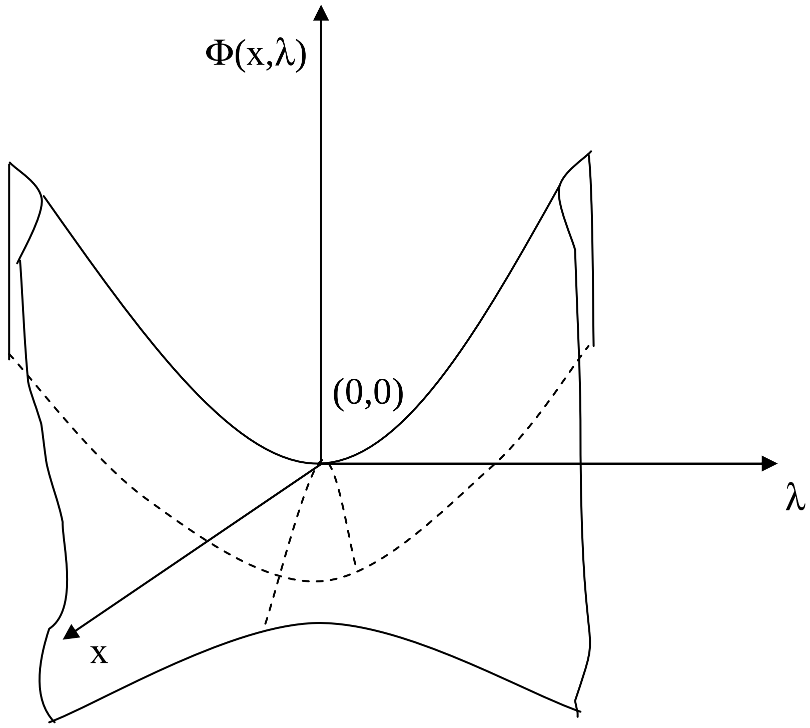

Рисунок 7.3 - Графік функції $\Phi(x,\lambda)$

Центральне місце в теорії нелінійного програмування займає теорема Куна-Таккера, що зв'язує розв’язання ЗНП з наявністю сідлової точки у відповідної функції Лагранжа.

Теорема 7.1 (достатня умова екстремуму).
Якщо (x^*, λ^*) - сідлова точка функції Лагранжа в області $x \in X \supseteq D$, $\lambda \geq 0$, то x^* є оптимальним планом задачі (7.11), причому справедливим є так зване правило нетвердості, що доповнює (умова Слейтера):

$$\sum_{i=1}^{m} \lambda_i^* g_i(x^*) = 0. \tag{7.14}$$

Доказ.

Скористаємося визначенням сідлової точки та запишемо

$$f(x) - \sum_{i=1}^{m} \lambda_i^* g_i(x) \leq f(x^*) - \sum_{i=1}^{m} \lambda_i^* g_i(x^*) \leq f(x^*) - \sum_{i=1}^{m} \lambda_i g_i(x^*) \tag{7.15}$$

при всіх $x \in X$, $\lambda \geq 0$. З другої нерівності у (7.15) випливає, що

$$\sum_{i=1}^{m} \lambda_i g_i(x^*) \leq \sum_{i=1}^{m} \lambda_i^* g_i(x^*) \text{ для } \lambda \geq 0. \tag{7.16}$$

Проте (7.16) може мати місце тільки тоді, коли $g(x^*) \leq 0$ при всіх $i = \overline{1, m}$. Дійсно, якщо існує таке *k*, що $g_k(x^*) > 0$, то поклавши $\lambda_i = 0$ для всіх $i \neq k$ і обравши досить велике $\lambda_k > 0$, можна домогтися того, що значення $\sum_{i=1}^{m} \lambda_i g_i(x^*) = \lambda_k g_k(x^*)$ виявиться більшим за постійний вираз $\sum_{i=1}^{m} \lambda_i^* g_i(x^*)$.

З того, що для всіх $i = \overline{1, m}$ виконуються нерівності $g_i(x^*) \leq 0$, випливає, що x^* є припустимим планом задачі (7.11).

Якщо до лівої частини нерівності (7.16) підставити значення $\lambda_i = 0$ для всіх $i = \overline{1, m}$, то одержимо

$$\sum_{i=1}^{m} \lambda_i^* g_i(x^*) \geq 0.$$

Разом з тим з того що, $g_i(x^*) \leq 0$ і $\lambda_i^* \geq 0$ випливає оцінка

$$\sum_{i=1}^{m} \lambda_i^* g_i(x^*) \leq 0.$$

Сумісний розгляд останніх двох нерівностей призводить до правила нетвердості, що доповнює, у точці x^*:

$$\sum_{i=1}^{m} \lambda_i^* g_i(x^*) = 0.$$

Тоді на підставі лівої частини нерівності сідлової точки (7.15) маємо, що для всіх $x \in X$ (у тому числі і для $x \in D$)

$$f(x^*) \geq f(x) - \sum_{i=1}^{m} \lambda_i^* g_i(x).$$

За умовою ЗНП для будь-яких $x \in D$ вірні нерівності $g_i(x^*) \le 0$, що в сполученні з умовою $\lambda_i^* = 0$ дозволяє записати

$$\sum_{i=1}^{m} \lambda_i^* g_i(x) \le 0.$$

Виходить,

$$f(x) - \sum_{i=1}^{m} \lambda_i^* g_i(x) \ge f(x).$$

Остаточно одержуємо, що для будь-яких $x \in D$ справедливе співвідношення $f(x^*) \ge f(x)$, тобто x^* - оптимальний план задачі (7.11).

Значення теореми Куна-Таккера полягає в тому, що вона дозволяє зв'язати процес розв'язання оптимізаційної задачі з пошуком сідлових точок функції Лагранжа, тобто з максимізацією цієї функції за x і мінімізацією за λ.

Нехай $F(x)$ є функцією, що ставить у відповідність кожному значенню x мінімальне значення функції $\Phi(x,\lambda)$ за λ:

$$F(x) = \min_{\lambda \ge 0} \Phi(x, \lambda)$$

і за аналогією

$$G(\lambda) = \max_{x \in X} \Phi(x, \lambda).$$

Розглянемо задачу відшукання максимуму функції $F(x)$

$$F(x) = \min_{\lambda \ge 0} \Phi(x, \lambda) \to \max, \quad x \in X \tag{7.17}$$

і задачу мінімізації $G(\lambda)$

$$G(\lambda) = \max_{x \in X} \Phi(x, \lambda) \to \min, \quad \lambda \ge 0. \tag{7.18}$$

Очевидно, що

$$F(x) = \min_{\lambda \ge 0} \Phi(x, \lambda) = \begin{cases} f(x), \ x \in D, \\ \infty, \quad x \notin D \end{cases}.$$

Звідси випливає, що максимум $F(x)$ знаходиться в припустимій області D і збігається з максимумом цільової функції $f(x)$ задачі (7.11):

$$\max_{x \in X} F(x) = \max_{x \in X} \min_{\lambda \ge 0} \Phi(x, \lambda) = \max_{x \in D} f(x).$$

Таким чином, задача (7.17), у певному змісті, рівносильна (7.11). Аналогічні висновки можна отримати і для (7.18). Задачі (7.17) і (7.18) утворюють двоїсту пару. Дане відношення є узагальненням відношення подвійності для задач лінійного програмування. Відповідно, за певних умов пара двоїстих задач нелінійного програмування має властивості, що аналогічні властивостям двоїстих лінійних задач. Зокрема, при будь-яких $x \in X$, $\lambda \ge 0$

$$F(x) \le G(\lambda). \tag{7.19}$$

Умова (7.19) знаходить широке застосування при побудові оцінок в ітеративних методах розв’язання оптимізаційних задач. Наприклад, якщо є можливість приблизно вирішити пряму та двоїсту задачі і одержати послідовності наближень, то за допомогою нерівностей виду

$$f(x^{(k)}) \le f(x^*) \le G(\lambda^{(k)})$$

можна визначити момент зупинки обчислювальної процедури.

Відзначимо, що можливим є варіант виведення виразів для цільових функцій і обмежень пари двоїстих задач лінійного програмування із загального визначення відношення подвійності для нелінійних задач. Також відзначимо, що в процесі формування нелінійних двоїстих задач існує велика неоднозначність: їхній вид можна варіювати, включаючи до множини X частину обмежень $g_i(x) \le 0$.

7.5 Деякі методи розв'язання задач НЛП

7.5.1 Градієнтні методи розв'язання задач безумовної оптимізації

Провідне місце серед прямих методів розв'язання екстремальних задач займає градієнтний метод (точніше, сімейство градієнтних методів) пошуку стаціонарних точок диференційованої функції. Нагадаємо, що стаціонарною називається точка, у якій $\nabla f(x)=0$ і яка відповідно до необхідної умови оптимальності може мати місце наявність локального екстремуму. Таким чином, застосовуючи градієнтний метод, знаходять множину точок локальних максимумів (або мінімумів), серед яких визначають максимум (або мінімум) глобальний.

Ідея даного методу ґрунтується на тому, що градієнт функції вказує напрямок її найбільш швидкого зростання в оточенні тієї точки, в якій він обчислений. Тому, якщо з певної поточної точки $x^{(1)}$ переміщатися в напрямку вектору $\nabla f(x^{(1)})$, функція f зростатиме, принаймні, у певному оточенні $x^{(1)}$. Отже, для точки $x^{(2)} = x^{(1)} + k\nabla f(x^{(1)})$, ($k>0$), що лежить у такому оточенні, справедлива нерівність $f(x^{(1)}) \le f(x^{(2)})$. Продовжуючи цей процес, ми поступово наближатимемося до точки певного локального максимуму.

Проте як тільки визначається напрямок руху, відразу ж встає питання, як далеко треба рухатися в цьому напрямку або, інакше кажучи, виникає проблема вибору кроку r у рекурентній формулі

$$x^{(k+1)}=x^{(k)}+r\nabla f(x^{(k)}), \qquad (7.20)$$

що задає послідовність точок, які прагнуть до точки максимуму.

Залежно від способу її розв'язання розрізняють різні варіанти градієнтного методу. Зупинимося на найбільш відомих з них.

Метод найшвидшого спуску. Назва методу пов'язана з мінімізацію цільової функції. Проте, за традицією така назва використовується і при розв'язанні задачі на максимум.

Нехай $f(x) = f(x_1, x_2, .., x_n)$ - диференційована функція, а $x^{(k)}=(x_1^{(k)}, x_2^{(k)}, ..., x_n^{(k)})$ - певна поточна точка. Будь-яких загальних рекомендацій щодо вибору вихідної точки (початкового наближення) $x^{(0)}$ не існує, але за можливістю вона має знаходитись близько від шуканого оптимального плану x^*. Якщо $x^{(k)}$ - нестаціонарна точка, то під час руху в напрямку $\nabla f(x^{(k)})$ функція $f(x)$ на певному інтервалі обов'язково зростатиме. Звідси виникає необхідність такого вибору кроку, щоб рух у зазначеному напрямку тривав доти, поки зростання не припиниться. Виразимо залежність значення $f(x)$ від крокового множника $r > 0$, покладаючи $x = x^{(k)} + r\nabla f(x^{(k)})$

$$f(x) = f(x^{(k)} + r\nabla f(x^{(k)})) = \varphi(r), \qquad (7.21)$$

або в координатній формі,

$$\varphi(r) = f\left(x_1^{(k)} + r\frac{\partial f(x^{(k)})}{\partial x_1}, \dots, x_n^{(k)} + r\frac{\partial f(x^{(k)})}{\partial x_n}\right). \qquad (7.22)$$

Щоб домогтися найбільшого з можливих значень f під час руху у напрямку $\nabla f(x^{(k)})$, потрібно вибрати таке значення r, що максимізує функцію $\varphi(r)$. Для обчислення r використовується необхідна умова екстремуму $\frac{d\varphi(r)}{dr} = 0$. Помітимо, що якщо для будь-якого r>0 $\frac{d\varphi(r)}{dr} > 0$, то функція $f(x)$ не обмежена зверху (тобто не має максимуму). У протилежному випадку, на підставі (7.22) одержуємо

$$\frac{d\varphi(r)}{dr} = \frac{\partial f(x)}{\partial x_1} \times \frac{dx_1}{dr} + \dots + \frac{\partial f(x)}{\partial x_n} \times \frac{dx_n}{dr}, \qquad (7.23)$$

що, у свою чергу, дає

$$\frac{d\varphi(r)}{dr} = \frac{\partial f(x)}{\partial x_1} \times \frac{\partial f(x^{(k)})}{\partial x_1} + \dots + \frac{\partial f(x)}{\partial x_n} \times \frac{\partial f(x^{(k)})}{\partial x_n} = \nabla f(x) \nabla f(x^{(k)}). \qquad (7.24)$$

Якщо вважати, що наступна точка $x^{(k+1)}$ відповідає оптимальному значенню, то в ній має виконуватися умова $\frac{d\varphi(r)}{dr} = 0$, і r треба знаходити з умови $\nabla f(x^{(k+1)}) \nabla f(x^{(k)}) = 0$ або

$$\nabla f(x^{(k)} + r\nabla f(x^{(k)})) \nabla f(x^{(k)}) = 0. \qquad (7.25)$$

Умова (7.25) означає рівність нулю скалярного добутку градієнтів функції f в точках $x^{(k+1)}$ і $x^{(k)}$. Геометрично її можна інтерпретувати як перпендикулярність векторів градієнтів функції f у зазначених точках. Продовжуючи геометричну інтерпретацію методу найшвидшого спуску, відзначимо, що в точці $x^{(k+1)}$ вектор $\nabla f(x^{(k+1)})$, будучи градієнтом, перпендикулярний до лінії рівня, що проходить через дану точку. Відповідно вектор $\nabla f(x^{(k)})$ є дотичним до цієї лінії. Таким чином, рух у напрямку градієнту $\nabla f(x^{(k)})$ треба продовжувати доти, поки він перетинає лінії рівня функції, що оптимізується.

Після того як точку $x^{(k+1)}$ знайдено, вона стає поточною для чергової ітерації. На практиці ознакою досягнення стаціонарної точки служить досить мала зміна координат точок, розглянутих на послідовних ітераціях. Одночасно з цим координати вектору $\nabla f(x^{(k)})$ повинні бути близькі до нуля.

Метод дроблення кроку. Для знаходження кроку r у методі найшвидшого спуску потрібно вирішити рівняння (7.25), що може виявитися досить складним. Тому часто обмежуються «підбором» такого значення r, що $\varphi(r) > \varphi(0)$. Для цього задаються певним початковим значенням r_1

(наприклад, $r_1=1$) і перевіряють умову $\varphi(r_1) > \varphi(0)$. Якщо вона не виконується, то покладають

$$r_2 = \frac{r_1}{2}$$

і т. д. доти, поки не вдається знайти підходящий крок, з яким переходять до наступної точки $x^{(k+1)}$. Критерій завершення алгоритму буде таким самим, як і в методі найшвидшого спуску.

7.5.2 Квадратичне програмування (КП)

До задач квадратичного програмування належить спеціальний клас задач нелінійного програмування, для яких цільова функція $f(x)$ квадратична, а всі обмеження - лінійні. Застосувавши до цієї задачі теорему Куна-Таккера, одержують умову для оптимального розв'язку у вигляді системи лінійних рівнянь, вирішити які можна симплекс-методом. У матричному вигляді задача квадратичного програмування формулюється в такий спосіб:

Максимізувати (мінімізувати) цільову функцію $f = \overline{C}\,\overline{X} + \overline{X}^T\,\overline{D}\,\overline{X}$ при обмеженнях

$$\overline{A}\,\overline{X} \le b, \quad \overline{X} \ge 0, \tag{7.26}$$

де

$$\overline{X} = (x_1, x_2, ..., x_n)^T,$$

$$\overline{C} = (c_1, c_2, ..., c_n),$$

$$b=(b_1, b_2, \dots, b_m)^T,$$

$$\overline{A} = \begin{pmatrix} a_{11}\, a_{12} \dots a_{1n} \\ a_{21}\, a_{22} \dots a_{2n} \\ \vdots \quad \vdots \quad \dots \vdots \\ a_{m1}\, a_{m2} \dots a_{mn} \end{pmatrix},$$

$$\overline{D} = \begin{pmatrix} d_{11}\, d_{12} \dots d_{1n} \\ d_{21}\, d_{22} \dots d_{2n} \\ \vdots \quad \vdots \quad \dots \vdots \\ d_{n1}\, d_{n2} \dots d_{nn} \end{pmatrix}.$$

Функція $\overline{X}^T\,\overline{D}\,\overline{X}$, де D – симетрична матриця, що є квадратичною формою (7.9). Матриця D є від'ємно визначеною в задачі максимізації і додатно визначеною – у задачі мінімізації. Це означає, що функція f є строго опуклою за змінними $\overline{X}$ у задачі мінімізації і строго ввігнутою – у задачі максимізації. Обмеження задачі є лінійними, що гарантує опуклість області припустимих розв'язків.

Складемо функцію Лагранжа для задачі (7.26)

$$\Phi(\overline{X}, \overline{\lambda}) = \overline{C}\,\overline{X} + \overline{X}^T\,\overline{D}\,\overline{X} + \overline{\lambda}(b - \overline{A}\,\overline{X})\,. \tag{7.27}$$

Відповідно до теореми Куна-Таккера для існування сідлової точки необхідно і достатньо, щоб похідні функції Лагранжа за x_j були менші за нуль, а за λ_i – перевищували б нуль, маємо:

$$\frac{\partial\Phi(\overline{X},\overline{\lambda})}{\partial x_j} = \overline{C} + 2\overline{D}\overline{X} - \overline{\lambda}\overline{A} \le 0, \qquad (7.28)$$

причому, якщо $\frac{\partial\Phi}{\partial x_j} < 0$, відповідні x_j дорівнюють нулю;

$$\frac{\partial\Phi(\overline{X},\overline{\lambda})}{\partial \lambda_i} = b - \overline{A}\overline{X} \ge 0, \qquad (7.29)$$

якщо $\frac{\partial\Phi}{\partial \lambda_i} > 0$, відповідні λ_i дорівнюють нулю.

Щоб нерівності (7.28) і (7.29) призвести до рівностей, уведемо два допоміжних вектори: $\overline{V} = (v_1, v_2, \dots, v_n) \ge 0$ і $\overline{S} = (s_1, s_2, \dots, s_m) \ge 0$, причому виберемо v_j>0, якщо $\frac{\partial\Phi}{\partial x_j} < 0$ і v_j=0, якщо $\frac{\partial\Phi}{\partial x_j} = 0$, а також λ_i>0, якщо $\frac{\partial\Phi}{\partial \lambda_i} > 0$ і λ_i=0, якщо $\frac{\partial\Phi}{\partial \lambda_i} = 0$, одержимо

$$\overline{C} + 2\overline{D}\overline{X} - \overline{\lambda}\overline{A} + \overline{V} = 0, \qquad (7.30)$$

$$b - \overline{A}\overline{X} - \overline{S} = 0. \qquad (7.31)$$

Порівнявши компоненти векторів $\overline{X}$ і $\overline{V}$, а також $\overline{\lambda}$ і $\overline{S}$, одержимо дві умови нетвердості, що доповнюють

$$\overline{X}^T \overline{V} = 0 \text{ і } \overline{\lambda}^T \overline{S} = 0. \qquad (7.32)$$

З умов (7.32) випливає, що хоча б n змінних з $\overline{X}$ і $\overline{V}$, а також m змінних з $\overline{\lambda}$ і $\overline{S}$ обертаються на нуль. Якщо існує оптимальний розв'язок задачі (7.26), то він є одним з базисних розв'язків системи (7.30), (7.31). Для знаходження припустимого базисного розв'язку використовується симплекс-метод.

Розглянемо приклад. Знайти

$$max\ f(x_1,x_2)=\max(10x_1+20x_2+x_1x_2-2x_1^2-2x_2^2)$$

за обмежень:

$$8-x_2 \ge 0,$$
$$10-x_1-x_2 \ge 0,$$
$$x_1 \ge 0;\ x_2 \ge 0.$$

Можна показати, що функція $f(x_1,x_2)$ є ввігнутою функцією. Складемо функцію Лагранжа

$$\Phi(x_1,x_2,\lambda_1,\lambda_2) = 10x_1+20x_2+x_1x_2-2x_1^2-2x_2^2+\lambda_1(8-x_2)+\lambda_2(10-x_1-x_2).$$

Застосувавши теорему Куна-Таккера, одержимо умови існування сідлової точки:

$$\frac{\partial \Phi}{\partial x_1} = 10 + x_2 - 4x_1 + \lambda_2 \le 0,$$

$$\frac{\partial \Phi}{\partial x_2} = 20 + x_1 - 4x_2 - \lambda_1 - \lambda_2 \le 0,$$

$$\frac{\partial \Phi}{\partial \lambda_1} = 8 - x_2 \ge 0,$$

$$\frac{\partial \Phi}{\partial \lambda_2} = 10 - x_1 - x_2 \ge 0.$$

Одержали систему лінійних нерівностей:

$$\begin{cases} 10 + x_2 - 4x_1 + \lambda_2 \le 0, \\ 20 + x_1 - 4x_2 - \lambda_1 - \lambda_2 \le 0, \\ 8 - x_2 \ge 0, \\ 10 - x_1 - x_2 \ge 0. \end{cases}$$

Уведемо вільні змінні $\overline{V}$ та $\overline{S}$, що обертають систему нерівностей на систему лінійних рівнянь, вирішивши яку симплекс-методом, знайдемо оптимум

$$\begin{cases} 4x_1 - x_2 - \lambda_2 - v_1 = 10, \\ -x_1 + 4x_2 + \lambda_1 + \lambda_2 - v_2 = 20, \\ x_2 + s_1 = 8, \\ x_1 + x_2 + s_2 = 10. \end{cases}$$

Для знаходження припустимого базисного плану отриманої задачі скористаємося методом мінімізації нев'язань. Для цього побудуємо допоміжну задачу, увівши до обмежень 1 і 2 фіктивні змінні x_3 і x_4, які входять у цільову функцію допоміжної задачі з коефіцієнтами M (а інші змінні – з нульовими коефіцієнтами). Таким чином, допоміжна задача має вигляд

$$f'(x)=Mx_3+Mx_4,$$

$$\begin{cases} 4x_1 - x_2 - \lambda_2 - v_1 + x_3 = 10, \\ -x_1 + 4x_2 + \lambda_1 + \lambda_2 - v_2 + x_4 = 20, \\ x_2 + s_1 = 8, \\ x_1 + x_2 + s_2 = 10. \end{cases}$$

Складемо симплекс-таблицю 7.1.

Таблиця 7.1 – Симплекс-таблиця допоміжної задачі

Базис	$C_{баз}$	C_j	0	0	0	0	0	0	M	M	0	0
		B	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}
A_7	M	10	4	-1	0	-1	-1	0	1	0	0	0
A_8	M	20	-1	4	1	1	0	-1	0	1	0	0
A_9	0	8	0	1	0	0	0	0	0	0	1	0
A_{10}	0	10	1	1	0	0	0	0	0	0	0	1
	L_j	30M	3M	3M	M	0	-M	-M	M	M	0	0
	Δ_j		3M	3M	M	0	-M	-M	0	0	0	0

Оскільки в першу чергу необхідно мінімізувати нев'язання, виведемо з базису вектор $\overline{A}_7$ і введемо вектор $\overline{A}_1$ (тим самим змінна x_3 стане рівною нулю, а змінна x_1 увійде до базису). Складемо нову симплекс-таблицю 7.2.

Таблиця 7.2 – Перша ітерація

Базис	$C_{баз}$	C_j	0	0	0	0	0	0	M	M	0	0
		B	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}
A_1	0	10/4	1	-1/4	0	-1/4	-1/4	0	1/4	0	0	0
A_8	M	90/4	0	15/4	1	3/4	-1/4	-1	1/4	1	0	0
A_9	0	8	0	1	0	0	0	0	0	0	1	0
A_{10}	0	30/4	0	5/4	0	1/4	1/4	0	-1/4	0	0	1
	L_j	$90M/4$	0	$15M/4$	M	$3M/4$	$-M/4$	$-M$	$M/4$	M	0	0
	Δ_j		0	$15M/4$	M	$3M/4$	$-M/4$	$-M$	$3M/4$	0	0	0

Тепер виведемо з базису вектор $\overline{A}_8$ і введемо $\overline{A}_2$ (табл. 7.3).

Таблиця 7.3 – Друга ітерація

Базис	$C_{баз}$	C_j	0	0	0	0	0	0	M	M	0	0
		B	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}
A_1	0	4	1	0	1/15	-1/5	-4/15	-1/15	4/15	1/15	0	0
A_2	0	6	0	1	4/15	3/15	-1/15	-4/15	1/15	4/15	0	0
A_9	0	2	0	0	-4/15	-1/5	1/15	4/15	-1/15	-4/15	1	0
A_{10}	0	0	0	0	-1/3	0	1/3	1/3	-1/3	-1/3	0	1
	L_j	0	0	0	0	0	0	0	0	0	0	0
	Δ_j	0	0	0	0	0	0	0	$-M$	$-M$	0	0

Таким чином, отриманий припустимий базисний план вихідної задачі:

$$x_1=4,\ x_2=6,\ \lambda_1=0,\ \lambda_2=0,\ v_1=0,\ v_2=0,\ s_1=2,\ s_2=0.$$

Перевіримо виконання умов нетвердості, що доповнюють (7.32) $\overline{X}^T\overline{V}=0$ і $\overline{\lambda}^T\overline{S}=0$:

$$x_1 * v_1=0,\ x_2 * v_2=0,\ \lambda_1 * s_1=0,\ \lambda_2 * s_2=0.$$

Оскільки умови нетвердості, що доповнюють, виконуються, отриманий план є оптимальним. Оптимальне значення цільової функції

$$f(x_1,x_2)= 10*4+20*6+4*6-2*4^2-2*6^2= 80.$$

Контрольні запитання

1. За яких умов оптимізаційна задача може бути віднесена до класу нелінійних?
2. Наведіть приклад економічної моделі, що збігається до задачі нелінійного програмування.
3. Перелічіть основні труднощі, що виникають у процесі розв’язання задачі нелінійного програмування.
4. Який зміст вкладається в поняття «умовна оптимізація»?
5. Для чого призначений метод множників Лагранжа та у чому він полягає?
6. Яка точка називається стаціонарною?
7. Які принципові етапи входять у градієнтні методи?
8. Для розв’язання яких задач призначений метод найшвидшого спуску та метод дроблення кроку?
9. Дайте визначення опуклої (увігнутої) функції.
10. Сформулюйте достатню умову опуклості (увігнутості) функції.

11. У чому полягає специфіка задач опуклого програмування?
12. Дайте визначення сідлової точки. Наведіть приклад функції, що має сідлову точку.
13. Сформулюйте необхідну й достатню умови теореми Куна-Таккера. Яке значення вони мають для розв’язання задач нелінійного програмування?
14. У чому полягає умова регулярності Слейтера? Поясните її зміст.
15. Наведіть приклад пари двоїстих задач нелінійного програмування.
16. Які властивості пари нелінійних двоїстих задач можуть бути застосовані для їхнього розв’язання?

ТЕМА 8
ДИНАМІЧНЕ ПРОГРАМУВАННЯ

8.1 Загальна схема методів динамічного програмування

Деякі оптимізаційні задачі мають специфічні особливості, які дозволяють звести їхнє розв’язання до розгляду ряду більш простих підзадач. У результаті глобальна оптимізація певної функції збігається до поетапної оптимізації проміжних цільових функцій. У динамічному програмуванні розглядають методи, що дозволяють шляхом поетапної (багатокрокової) оптимізації одержати загальний оптимум.

Методами динамічного програмування оптимізують роботу керованих систем, ефективність яких оцінюється адитивною, або мультиплікативною цільовою функцією. Адитивною називається така функція кількох змінних $f(x_1, x_2, \ldots, x_n)$, значення якої обчислюється як сума функцій f_j, що залежать тільки від однієї змінної x_j:

$$f(x_1, x_2, \ldots, x_n) = \sum_{j=1}^{n} f_j(x_j)\,. \tag{8.1}$$

Доданки адитивної цільової функції відповідають ефективності розв’язків, прийнятих на окремих етапах керованого процесу. За аналогією, мультиплікативна функція розпадається на добуток додатних функцій різних змінних:

$$f(x_1, x_2, \ldots, x_n) = \prod_{j=1}^{n} f_j(x_j)\,. \tag{8.2}$$

Оскільки логарифм функції (8.2) є адитивною функцією, досить обмежитися розглядом функцій виду (8.1).

Обчислення в динамічному програмуванні виконуються рекурентно (рекурентний означає зворотний), так що оптимальний розв’язок однієї підзадачі використовується як вихідні дані для наступної. Вирішивши останню підзадачу, одержують оптимальний розв’язок вихідної задачі - глобальний оптимум.

Розглянемо приклад. Нехай необхідно обрати найкоротший шлях між двома містами. Сітка доріг (рис. 8.1) утворює можливі маршрути між вихідним містом, що знаходиться у вузлі 1, і кінцевим пунктом, що знаходиться у вузлі 7. Маршрути проходять через проміжні міста, позначені вузлами з номерами 2-6.

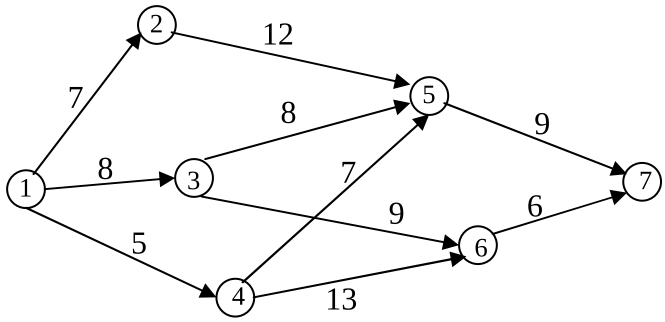

Рисунок 8.1 - Сітка доріг, що утворює можливі маршрути

Дану задачу можна вирішити шляхом перебору всіх маршрутів між вузлами 1 і 7. Проте у великій сітці такий підхід є неефективним. Для розв’язання задачі за методом динамічного програмування розділимо її на три етапи (рис. 8.2) і виконаємо обчислення для кожного етапу окремо.

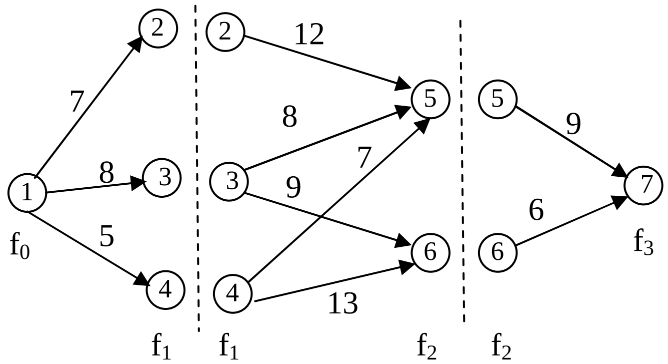

Рисунок 8.2 - Декомпозиція задачі

Обчислимо найкоротші відстані до всіх вузлів першого етапу з наступним використанням цих відстаней як вихідних даних для другого етапу. На першому етапі кожний з вузлів 2, 3 і 4 зв'язаний з початковим вузлом 1 єдиною дугою, маємо

Найкоротший шлях (1-2) дорівнює f_1=7 км;

Найкоротший шлях (1-3) дорівнює f_1=8 км;

Найкоротший шлях (1-4) дорівнює f_1=5 км.

На другому етапі обчислимо найкоротші відстані до вузлів 5 і 6. Розглянемо вузол 5. До нього ведуть три маршрути (2, 5), (3, 5) і (4, 5). Найкоротша відстань до вузлу 5 з урахуванням відстаней до вузлів 2, 3 і 4 визначиться в такий спосіб

$$f_2 = \min_{i=2,3,4}\left\{ \begin{matrix} f_1 + \textit{відстань від} \\ \textit{вузлу } i \textit{ до вузлу } 5 \end{matrix} \right\} = \min\begin{pmatrix} 7+12=19 \\ 8+8=16 \\ 5+7=12 \end{pmatrix} = 12 \; (\textit{з вузлу } 4).$$

Аналогічно для вузлу 6 одержимо

$$f_2 = \min_{i=2,3,4}\left\{\begin{matrix} f_1 + \textit{відстань от} \\ \textit{вузлу } i \textit{ до вузлу } 6 \end{matrix}\right\}\textit{від} = \min\begin{pmatrix} 8+9=17 \\ 5+13=18 \end{pmatrix} = 17 \ (\textit{з вузлу } 3)\,.$$

Звернемося до третього етапу. Кінцевий вузол 7 можна досягти як з вузлу 5 так і з вузлу 6. Використовуючи результати другого етапу, одержимо

$$f_3 = \min_{i=5,6}\left\{\begin{matrix} f_2 + \textit{відстань від} \\ \textit{вузлу } i \textit{ до вузлу } 7 \end{matrix}\right\} = \min\begin{pmatrix} 12+9=21 \\ 17+6=23 \end{pmatrix} = 21 \ (\textit{з вузлу } 5)\,.$$

Остаточно, найкоротший шлях до вузлу 7 становить 21 км, оптимальний маршрут - послідовність вузлів 1, 4, 5, 7.

Рекурентні обчислення динамічного програмування виражаються математично в такий спосіб. Нехай $f_i(x_i)$ – найкоротша відстань до вузлу x_i на етапі i, $d(x_{i-1}, x_i)$ – відстань від вузлу x_{i-1} до вузлу x_i, тоді f_i обчислюється на підставі рекурентного виразу

$$f_i(x_i) = \min\{d(x_{i-1}, x_i) + f_{i-1}(x_{i-1})\}, \quad i = \overline{1,n}\,. \qquad (8.1)$$

Рівняння (8.1) показує, що функція цілі $f_i(x_i)$ на і-му етапі має бути вираженою як функція попереднього етапу. У термінології динамічного програмування x_i називається **станом системи** на етапі i. Стан системи на етапі i – це інформація, що зв'язує етапи один з одним, при цьому оптимальні розв’язки для етапів, що залишилися, можуть прийматися без повторної перевірки того, як були отримані розв’язки на попередніх етапах. Таке визначення стану системи дозволяє розглядати кожний етап окремо і гарантує, що розв’язок є припустимим на кожному етапі. Таким чином, можна сформулювати **принцип оптимальності**:

> На кожному етапі оптимальна стратегія визначається незалежно від стратегій, що використовані на попередніх етапах. (Оптимальна стратегія управління не залежить від передісторії системи, а залежить від стану системи в сучасний момент часу й цілі управління).

Послідовність обчислень, що використована в прикладі, називається **алгоритмом прямого прогону**. Цю саму задачу можна вирішити за допомогою **алгоритму зворотного прогону**, за яким обчислення проводять від третього етапу до першого. Алгоритми прямого та зворотного прогону призведуть до того самого результату, але на практиці частіше використовують алгоритм зворотного прогону, тому що він ефективніший з обчислювальної точки зору.

Для алгоритму зворотного прогону рекурентне рівняння (8.1) має вигляд

$$f_i(x_i) = \min\{d(x_i, x_{i+1}) + f_{i+1}(x_{i+1})\}, \quad i = \overline{1,n}\,. \qquad (8.2)$$

Вирішимо задачу, використовуючи алгоритм зворотного прогону.

Етап 3. Вузол 7 зв'язаний з вузлами 5 і 6 тільки одним маршрутом. Результати третього етапу занесемо в таблицю 8.1.

Таблиця 8.1 – Етап 3

	$d(x_3, x_4)$	**Оптимальний розв'язок**	
x_3	x_4=7	$f_3(x_3)$	x^*_4
5	9	9	7
6	6	6	7

Етап 2. Використовуючи значення $f_3(x_3)$, отримані на третьому етапі, порівняємо припустимі альтернативні розв'язки в таблиці 8.2.

Таблиця 8.2 – Етап 2

	$d(x_2, x_3)$+ $f_3(x_3)$		**Оптимальний розв'язок**	
x_2	x_3=5	x_3=6	$f_2(x_2)$	x^*_3
2	12+9=21	-	21	5
3	8+9=17	9+6=15	15	6
4	7+9=16	13+6=19	16	5

Оптимальний розв'язок другого етапу означає, що з вузлу 2 або 4 найкоротший шлях до вузлу 7 проходить через вузол 5, а з вузлу 3 - через вузол 6.

Етап 1. Використовуючи значення $f_2(x_2)$, отримані на другому етапі, порівняємо припустимі альтернативні розв'язки в таблиці 8.3.

Таблиця 8.3 – Етап 1

	$d(x_2, x_3)$+ $f_3(x_3)$			**Оптимальний розв'язок**	
x_1	x_2=2	x_2=3	x_2=4	$f_1(x_1)$	x^*_2
1	7+21=28	8+15=23	5+16=21	21	4

8.2 Основні типи задач і моделі динамічного програмування

Задача про завантаження. Це задача про раціональне завантаження судна, що має обмеження за обсягом або вантажопідйомністю. Кожний поміщений на судно вантаж приносить певний прибуток. Задача полягає у визначенні завантаження судна такими вантажами, які принесуть найбільший сумарний прибуток. Дана задача відома також як задача про ранець, у якої треба визначити найцінніші предмети, що підлягають завантаженню в ранець.

Нехай вантажопідйомність судна W предметів n найменувань, m_i – кількість предметів i-го найменування, що підлягають завантаженню, r_i – прибуток, що приносить один завантажений предмет i-го найменування, w_i – вага одного предмета i-го найменування. Формулювання задачі має вигляд:

Максимізувати $f=r_1m_1+r_2m_2+\ldots+r_nm_n$

за умови

$$w_1m_1+w_2m_2+\ldots+w_nm_n\leq W;$$
$$m_1, m_2,\ldots,m_n\geq 0 \quad \text{і цілі.}$$

Елементи моделі динамічного програмування визначаються в такий спосіб. Етап i ставиться у відповідність предмету i-го найменування; варіанти розв'язку на етапі i описуються кількістю m_i предметів i-го найменування, що підлягають завантаженню. Відповідний прибуток дорівнює $r_i m_i$. Значення m_i лежить у межах від 0 до $\frac{W}{w_i}$. Стан x_i на етапі i виражає сумарну вагу предметів, рішення про завантаження яких прийняті на етапах i.

Рекурентне рівняння має вигляд

$$f_i(x_i) = \max\{r_i m_i + f_{i+1}(x_i - w_i m_i)\}, \quad i = \overline{1,n}. \tag{8.3}$$

Задача про завантаження є типовим представником задачі розподілу ресурсів, у якій обмежений ресурс розподіляється між кінцевим числом видів економічної діяльності. При цьому ціллю є максимізація прибутку. У таких моделях станом на етапі i є сумарна кількість ресурсу, що розподіляється на відповідному етапі.

Розглянемо приклад. У 4-тонне судно завантажують предмети трьох найменувань. Дані про вагу одного предмета та прибуток, одержуваний від одного завантаженого предмета, наведені в таблиці 8.4.

Таблиця 8.4 – Вихідні дані

Предмет *i*	w_i	r_i
1	2	31
2	3	47
3	1	14

Етап 3. Вага, що може бути завантаженою на етапі 3 (предмет 3) може приймати одне із значень 0, 1, 2, 3, 4 (тому що вантажопідйомність судна становить 4 тонни). Основою для порівняння варіантів на даному етапі є співвідношення

$$f_3(x_3) = \max\{14 m_3\}, \quad \max\{m_3\} = \frac{4}{1} = 4.$$

Припустимі розв'язки для кожного значення x_3 порівняємо в таблиці 8.5.

Таблиця 8.5 – Етап 3

	14 m_3					**Оптимальний розв'язок**	
x_3	m_3=0	m_3=1	m_3=2	m_3=3	m_3=4	$f_3(x_3)$	m^*_3
0	0	-	-	-	-	0	0
1	0	14	-	-	-	14	1
2	0	14	28	-	-	28	2
3	0	14	28	42	-	42	3
4	0	14	28	42	56	56	4

Етап 2. $f_2(x_2) = \max\{47m_2 + f_3(x_2 - 3m_2)\}, \quad \max\{m_2\} = \frac{4}{3} = 1.$

Таблиця 8.6 – Етап 2

	47 m_2+f_3(x_2-3m_2)		**Оптимальний розв'язок**	
X_2	m_2=0	m_2=1	$f_2(x_2)$	m^*_2
0	0+0=0	-	0	0
1	0+14=14	-	14	0
2	0+28=28	-	28	0
3	0+42=42	47+0=47	47	1
4	0+56=56	47+14=61	61	1

Етап 1. $f_1(x_1) = \max\{31m_1 + f_2(x_1 - 2m_1)\}, \quad \max\{m_1\} = \frac{4}{2} = 2.$

Таблиця 8.7 – Етап 1

	31 m_3 + f_2(x_1-2m_1)			**Оптимальний розв'язок**	
X_1	m_1=0	m_1=1	m_1=2	$f_1(x_1)$	m^*_1
0	0+0=0	-	-	0	0
1	0+14=14	-	-	14	0
2	0+28=28	31+0=31	-	31	1
3	0+47=47	31+14=45	-	47	0
4	0+61=61	31+28=59	62+0=62	62	2

З умови W=4 випливає, що перший етап при x_1=4 дає оптимальний розв'язок, що означає, що будуть завантажені два предмети першого найменування. Це завантаження залишає для предмета другого найменування x_2= x_1-2m^*_1=4-2*2=0. Розв'язання на другому етапі при x_2=0 призводить до оптимального розв'язку m^*_2=0, що дає x_3= x_2-3m^*_2=0+3*0=0. Далі етап 3 при x_3=0 призводить до m^*_3=0. Таким чином, оптимальним розв'язком задачі є m^*_1=2, m^*_2=0 і m^*_3=0. Відповідний прибуток становитиме 62 грошові одиниці.

Задача планування робочої сили. При виконанні певних проектів число робітників, необхідне для реалізації будь-якого проекту, регулюється шляхом їхнього наймання та звільнення. Наймання та звільнення робітників пов'язане з додатковими витратами. Треба визначити, як має регулюватися чисельність робітників.

Припустимо, що проект виконуватиметься протягом n тижнів і мінімальна потреба в робочій силі протягом i-го тижня складе b_i робітників. Проте залежно від вартісних показників може бути вигіднішим відхилення чисельності робітників як в один так і в інший бік від мінімальної потреби. Якщо x_i - кількість робітників протягом i-го тижня, то можливі витрати двох видів: $C_1(x_i-b_i)$ - витрати пов'язані з необхідністю надлишку x_i-b_i робочої сили та

$C_2(x_i\text{-}x_{i-1})$ - витрати, пов'язані з необхідністю додаткового наймання x_i-x_{i-1} робітників.

Елементи моделі динамічного програмування в цій задачі визначаються в такий спосіб. Етап *i* представляється порядковим номером тижня *i*, $i=\overline{1,n}$. Варіантами розв'язку на *i*-му етапі є значення x_i - кількість робітників протягом *i*-го тижня. Станом на *i*-му етапі є x_{i-1} - кількість робітників протягом (*i*-1)-го тижня (етапу).

Рекурентне рівняння динамічного програмування має вигляд

$$f_i(x_{i-1})=\min\{C_1(x_i-b_i)+C_2(x_i-x_{i-1})+f_{i+1}(x_i)\},\quad i=\overline{1,n}. \tag{8.4}$$

Обчислення починаються з етапу *n* при $x_n=b_n$ і закінчуються на етапі 1.

Задача заміни устаткування. Чим довше механізм експлуатується, тим вище витрати на його обслуговування й нижче його продуктивність. Коли строк експлуатації механізму досягає певного рівня, може виявитися вигіднішою його заміна. Задача заміни обладнання збігається до визначення оптимального строку експлуатації механізму.

Нехай на початку кожного року приймається рішення або про експлуатацію механізму ще один рік, або про заміну його новим. Позначимо *r*(t) і *c*(t) прибуток від експлуатації *t*-літнього механізму та витрати на його обслуговування протягом року. Нехай *s*(*t*) - вартість продажу механізму, що експлуатувався *t* років. Вартість придбання нового механізму залишається незмінною протягом всіх *n* років і дорівнює *I*.

Елементи моделі динамічного програмування визначаються в такий спосіб. Етап *i* представляється порядковим номером року *i*, $i=\overline{1,n}$. Варіантами розв'язку на *i*-му етапі є альтернативи: продовжити експлуатацію або замінити механізм на початку *i*-го року. Станом на *i*-му етапі є тривалість експлуатації *t* механізму до початку *i*-го року.

Нехай $f_i(t)$ – максимальний прибуток, одержуваний за роки від *i* до *n* за умови, що на початку *i*-го року є механізм *t*-літнього віку.

Рекурентне рівняння має такий вигляд

$$f_i(t)=\max\left\{\begin{matrix} r(t)-c(t)+f_{i+1}(t+1), & \text{якщо експлуатирувати механізм,} \\ r(0)+s(t)-I-c(0)+f_{i+1}(1), & \text{якщо замінити механізм.} \end{matrix}\right\}. \tag{8.5}$$

Задача інвестування. Припустимо, що на початку кожного з наступних *n* років необхідно зробити інвестиції P_1, P_2, …, P_n. Є можливість вкласти капітал у два банки, перший з яких виплачує річний складний відсоток r_1, а другий – r_2. Для заохочення депозитів обидва банки виплачують новим інвесторам премії у вигляді відсотку від вкладеної суми. Преміальні змінюються від року до року й для *i*-го року дорівнюють q_{i1} і q_{i2} у першому й другому банках відповідно. Їх виплачують наприкінці року, і вони можуть бути інвестовані в один з двох банків на наступний рік. Розміщений у банку внесок повинен знаходитися там до кінця розглянутого періоду. Необхідно розробити стратегію інвестицій на наступні *n* років.

Елементи моделі динамічного програмування визначаються в такий спосіб. Етап *i* подається порядковим номером року *i*, $i=\overline{1,n}$. Варіантами розв'язку на *i*-му етапі є суми I_i^1 і I_i^2 інвестицій у перший і другий банк відповідно. Станом x_i на i-му етапі є сума грошей на початок i-го року, які можна інвестувати.

Оскільки $I_i^2 = x_i = I_i^1$, то $x_i = P_i + q_{i-1,1} I_{i-1}^1 + q_{i-1,2}(x_{i-1} - I_{i-1}^1) = P_i + (q_{i-1,1} + q_{i-1,2}) I_{i-1}^1 + {}$ $+ q_{i-1,2} x_{i-1}$. Нехай $f_i(x_i)$ – оптимальна сума інвестицій для інтервалу від i-го до n-го року за умови, що на початку i-го року є грошова сума x_i. Далі позначимо через s_i суму, накопичену до кінця n-го року за умови, що I_i^1 і (x_i -I_i^1) – обсяги інвестицій протягом i-го року в перший і другий банк відповідно. Позначивши $\alpha_i = (I^1 + r_i)$, i=1,2, сформулюємо задачу:

Максимізувати $z = s_1 + s_2 + \ldots + s_n$, де

$$s_i = I_i^1 \alpha_1^{n+1-i} + (x_i - I_i^1)\alpha_2^{n+1-i} = \left(\alpha_1^{n+1-i} - \alpha_2^{n+1-i}\right) I_i^1 + \alpha_2^{n+1-i} x_i, \quad i = \overline{1, n-1},$$

$$s_n = (\alpha_1 + q_{n1} - \alpha_2 - q_{n2}) I_n^1 + (\alpha_2 + q_{n2}) x_n.$$

Оскільки преміальні за n-й рік є частиною накопиченої грошової суми, у вираз для s_n додані q_{n1} і q_{n2}. У цьому випадку рекурентне рівняння для зворотного прогону має вигляд

$$f_i(x_i) = \max\{s_i + f_{i+1}(x_{i+1})\}, \quad i = \overline{1, n-1}. \tag{8.6}$$

Задача управління запасами. Однією з найбільш відомих сфер використання методів динамічного програмування є така область, як теорія управління запасами. Її предметом є розробка та дослідження математичних моделей систем, що займають проміжне положення між джерелами (виробниками) тих або інших ресурсів і їхніми споживачами. При математичній формалізації процесів управління запасами дуже часто доводиться використовувати стрибкоподібні, недиференційовані та кусочно-безперервні функції. Як правило, це зумовлюється необхідністю урахування ефектів концентрації, фіксованих витрат і плати за замовлення. У зв'язку з цим одержувані задачі трудно піддаються аналітичному розв’язанню класичними методами, проте можуть бути успішно вирішені за допомогою апарату динамічного програмування. Найпростіші задачі характеризуються постійним у часі попитом, миттєвим поповненням запасу та відсутністю дефіциту. Розглянемо модель з витратами на оформлення замовлення, у якій передбачається, що дефіцит не припускається. Уведемо позначення: z_i - кількість замовленої продукції, D_i - потреба в продукції (попит), x_i - обсяг запасу на початок етапу i, K_i - витрати на оформлення замовлення, h_i - витрати на зберігання одиниці продукції, що переходить з етапу i в етап i+1.

Відповідна функція витрат для етапу і задається формулою

$$C_i(z_i) = \begin{cases} 0, & z_i = 0, \\ K_i + c_i(z_i), & z_i > 0, \end{cases}$$

де $c_i(z_i)$ - функція граничних витрат при заданому значенні z_i.

Оскільки дефіцит не припускається, задача управління запасами збігається до обчислення значень z_i, що мінімізують сумарні витрати, пов'язані з розміщенням замовлень, закупівлею та зберіганням продукції протягом n етапів. Витрати на зберігання на i-му етапі передбачаються пропорційними величині

$$x_{i+1} = x_i + z_i - D_i,$$

яка являє собою обсяг запасу, що переходить з етапу i в етап i+1.

Для рекурентного рівняння процедури прямого прогону стан на етапі і визначається як обсяг запасу x_{i+1} на кінець етапу

$$0 \le x_{i+1} \le D_{i+1} + \ldots + D_n.$$

Ця нерівність означає, що в граничному випадку запас x_{i+1} може задовольнити попит на всіх наступних етапах.

Нехай $f_i(x_{i+1})$ - мінімальні загальні витрати на етапах 1, 2, … i при заданій величині запасу x_{i+1} на кінець етапу i. Тоді рекурентне рівняння алгоритму прямого прогону можна записати у вигляді

$$f_1(x_2) = \min\{C_1(z_1) + h_1 x_2\},$$

$$f_i(x_{i+1}) = \min\{C_i(z_i) + h_i x_{i+1} + f_{i-1}(x_{i+1} + D_i - z_i)\}, \quad i = 2, n. \tag{8.7}$$

Контрольні запитання

1. Для розв'язання яких задач призначений метод динамічного програмування?
2. Поясніть, що являють собою адитивна та мультиплікативна функції.
3. У чому полягає сутність методу динамічного програмування?
4. Яким умовам повинна задовольняти задача, щоб для її розв'язання міг бути застосований алгоритм динамічного програмування?
5. Сформулюйте принцип оптимальності Беллмана, поясніть відсутність післядії.
6. Чим визначається напрямок розв'язання задачі в алгоритмах динамічного програмування?
7. Сформулюйте оптимізаційну модель для задачі про планування робочої сили.
8. Випишіть основне рекурентне співвідношення, використовуване при розв'язанні задачі про планування робочої сили.
9. Сформулюйте оптимізаційну модель для задачі про заміну обладнання.
10. Випишіть основне рекурентне співвідношення, використовуване при розв'язанні задачі про заміну встаткування.
11. З якими особливостями задач управління запасами зв'язане застосування при їхньому розв'язанні апарату динамічного програмування?
12. Який вигляд має цільова функція в динамічній задачі управління запасами?
13. Випишіть основне рекурентне співвідношення, використовуване при розв'язанні динамічної задачі управління запасами.

ТЕМА 9
ПРИЙНЯТТЯ РІШЕНЬ В УМОВАХ НЕВИЗНАЧЕНОСТІ ТА РИЗИКУ

9.1 Загальна математична постановка задачі стохастичного програмування (СП)

Економічні системи функціонують в умовах невизначеності, що обумовлює ризикованість прийнятих рішень. Прийняття рішень в умовах невизначеності й ризику вимагає визначення альтернативних дій, причому, досліджуване явище може характеризуватися заданим імовірнісним розподілом деяких параметрів. У такому випадку для прийняття рішень використовують методи стохастичного програмування. Суть цих методів полягає в тому, що рішення залежить не тільки від керованих змінних $X=(x_1, x_2, \ldots, x_n)$, а також від ряду випадкових некерованих параметрів $\omega=(\omega_1, \omega_2, \ldots, \omega_k)$. Предметом стохастичного програмування є екстремальні задачі, у яких параметри умов або складові розв'язку – випадкові величини. Особливістю задач стохастичного програмування є те, що основні труднощі виникають не стільки при розробці методів їх розв'язання, скільки при постановці задачі. Це пов’язано з тим, що постановка задачі істотно залежить від структури наявної інформації й повинна відбивати особливості ухвалення рішення в умовах невизначеності. Зокрема, якщо вектор X є детермінованим, то він не залежить від випадкових параметрів моделі, а інакше - треба враховувати його залежність від випадкових параметрів. Важливо також, що розуміють під максимізацією (мінімізацією) цільової функції: максимізація її абсолютних значень або максимізація її математичного сподівання або будь-якої іншої імовірнісної характеристики (наприклад, моди або середнього квадратичного відхилення). Треба вирішити також як повинні виконуватися обмеження: абсолютно для всіх ω або в середньому, або з малою імовірністю припускається порушення обмежень задачі.

Таким чином, постановка задачі стохастичного програмування вимагає урахування економічних особливостей досліджуваного явища, а також евристичного підходу до її формулювання.

При виборі цільової функції в задачах стохастичного програмування треба визначити, чи цікавить нас, у першу чергу, максимум математичного сподівання економічного критерію (наприклад, прибутку або рентабельності), або мінімум його дисперсії. Також як цільову функцію можна прийняти лінійну комбінацію математичного сподівання й дисперсії економічного показника або ймовірність перевищення ним деякого фіксованого значення. Помітимо, що для стохастичних задач актуальною є багатокритеріальна оптимізація.

Щодо обмежень, то вимога, щоб оптимальний розв'язок задовольняв їм за будь-яких значень випадкових параметрів ω, є занадто жорстким. Зазвичай припускають невиконання обмежень з певною досить малою імовірністю α. Таким чином, задачі стохастичного програмування мають вигляд

знайти $$\max M[f(x, \omega)] \qquad (9.1)$$

за обмежень

$$P\{g(x, \omega)\leq 0\}\geq 1-\alpha. \qquad (9.2)$$

У задачі (9.1)-(9.2) необхідно максимізувати середній сподіваний ефект за умови, що обмеження виконуються з імовірністю 1-α.

Або інакше

знайти $$\max \xi \qquad (9.3)$$

за обмежень

$$P\{f(x, \omega)\geq\xi, g(x, \omega)\leq 0\}\geq 1-\alpha. \qquad (9.4)$$

У задачі (9.3)-(9.4) додаткова вимога, щоб значення критерію ефективності було не менш ξ з імовірністю 1-α, а значення ξ було максимальним.

Оптимальні плани, отримані на підставі моделей (9.1)-(9.2) або (9.3)-(9.4) називають *M*-планами, тому що критерієм оптимальності є математичне сподівання $f(x, \omega)$. Якщо як критерій прийнято дисперсію функції $f(x, \omega)$, то оптимальний план називається *D*-планом. Іноді як критерій оптимальності приймають різницю $M[f(x, \omega)] - KD[f(x, \omega)]$, де K - відомий параметр.

9.2 Класифікація задач стохастичного програмування. Основні методи розв'язання

Постановки задач стохастичного програмування розрізняються за трьома ознаками: за характером рішень; за показником якості розв'язку; за способом розчленовування обмежень задачі.

Задачі з цільовою функцією виду $M[C\ X]$ називають *M*-моделями, задачі, в яких потрібно мінімізувати величину дисперсії, називають *D*-моделями, а стохастичні задачі, у яких максимізується імовірність P, прийнято називати *P*-моделями. До цієї ж групи моделей включають також і задачі, де потрібно мінімізувати або максимізувати поріг ξ.

До одноетапних задач стохастичного програмування належать задачі, в яких рішення приймаються на підставі відомих стохастичних характеристик розподілу випадкових параметрів умов задачі, що отримані до спостереження за реалізаціями поточних значень цих параметрів. При цьому повинне прийматися певне найкраще в середньостатистичному значенні рішення.

Постановка задач у стохастичному програмуванні істотно залежить від можливості уточнення стану економічного середовища при реалізації оптимального рішення. Для економічних систем розробляють стратегічні й тактичні плани. У стратегічних планах враховують всі можливі значення ω. У певний момент часу в результаті спостереження стан економічного середовища стає відомим, тоді розробляють тактичний план, тобто знаходять розв'язок $X(\omega)$ при заданому ω. У загальному випадку спостереження не повністю визначають стан економічного середовища, тому етапи прийняття рішень можуть чергуватися з етапами спостережень. У цьому випадку виникає *N*-етапна задача стохастичного програмування. Властивість плану адаптуватися до умов його реалізації, що постійно змінюються, є необхідним для ефективного розвитку й функціонування економічної системи. Програмну частину плану вибирають таким чином, щоб максимізувати сподівану корисність із урахуванням майбутньої адаптації. План-адаптація повинен бути оптимальним для конкретної економічної ситуації і при цьому зберігати стратегію розвитку системи в цілому.

Методи розв'язання стохастичних задач поділяють на дві групи - прямі та непрямі. Прямі методи використовують, якщо на підставі інформації про параметр ω можна побудувати функції $f(x, \omega)$ і $g_i(x, \omega)$. У непрямих методах стохастичну задачу приводять до задачі лінійного або нелінійного програмування, тобто розглядають детермінований аналог задачі стохастичного програмування.

Зокрема, імовірнісне динамічне програмування відрізняється від детермінованого тим, що стани та прибутки на кожному етапі є випадковими величинами. Моделі імовірнісного динамічного програмування виникають, наприклад, при розгляданні стохастичних моделей управління запасами. Одноетапні моделі управління запасами відбивають ситуацію, коли для задоволення попиту протягом певного періоду продукція замовляється тільки один раз. Багатоетапні моделі використовують при плануванні запасів на n періодів. Вони припускають, що попит у кожний період описується стаціонарною (незалежною від часу) щільністю імовірностей.

Апарат теорії ігор розрізняє стратегічні та статистичні ігри. В основі стратегічних ігор лежить припущення, що кожен з гравців діє активно, і їхні інтереси є протилежними. У теорії статистичних ігор одним з гравців є природа, тобто сукупність зовнішніх обставин, в умовах яких доводиться приймати рішення. Неминучою платою за спробу одержати рішення в умовах неповної інформації про закони природи є прийняття помилкових рішень. Теорія статистичних ігор дозволяє виробити таку стратегію відносно прийняття рішень, що хоча й не виключає можливості прийняття невірних рішень, але зводить до мінімуму пов'язані з цим небажані наслідки.

Розглянемо приклад. Нехай необхідно зробити запас n товарів у кількості $X=(x_1, x_2, \ldots, x_n)$, на які є випадковий попит $\omega=(\omega_1, \omega_2, \ldots, \omega_n)$. Нестача одиниці j-го товару обкладається штрафом c_j, тобто $C=(c_1, c_2, \ldots, c_n)$, а витрати на зберігання одиниці відповідного товару, що не вдалося збути, задані вектором $d=(d_1, d_2, \ldots, d_n)$. Функція збитку, що відповідає розв'язку x, має вигляд

$$f(x,\omega) = \sum_{j=1}^{n} \left\{ Cj \max\left(0, \omega_j - x_j\right) + d_j \max\left(0, x_j - \omega_j\right) \right\},$$

де $Cj \max\left(0, \omega_j - x_j\right)$ - штраф за незадоволення попиту на j-й вид товару;

$d_j \max\left(0, x_j - \omega_j\right)$ - витрати на зберігання j-го виду товару.

Для знаходження оптимального розв'язку даної задачі необхідно знати функцію розподілу випадкової величини ω. Оскільки така функція невідома, вважають, що ω розподілена рівномірно. При цьому необхідно враховувати, що дане припущення може призвести до ухвалення неправильного рішення.

9.3 Імітаційне моделювання

Методи імітаційного моделювання дозволяють зібрати необхідну інформацію про поведінку системи шляхом створення її комп'ютеризованої моделі. Імітаційне моделювання не вирішує оптимізаційних задач, а являє собою техніку оцінки значень функціональних характеристик системи, що моделюється. Методи імітаційного моделювання знаходять широке застосування в економічних і комерційних задачах, включаючи оцінку поведінки споживача, визначення цін, економічне прогнозування діяльності фірм, у соціальних задачах та ін.

Попередником сучасного імітаційного моделювання вважається метод Монте-Карло, основна ідея якого полягає у використанні вибірки випадкових чисел для одержання імовірнісних або детермінованих оцінок будь-яких величин. Імітація є випадковим експериментом, тому будь-який результат імітаційного моделювання піддається експериментальним помилкам і підлягає статистичній перевірці. Для будь-якого експерименту також важливим є питання, яким має бути обсяг вибірки n і число реалізацій досліджуваної випадкової величини N.

Відмінність сучасних імітаційних моделей від методу Монте-Карло полягає в тому, що імітаційна модель зазвичай пов'язана з вивченням реально існуючої системи, поведінка якої є функцією часу. Існує два типи імітаційних моделей. Безперервні моделі використовуються для систем, поведінка яких змінюється в часі безупинно. Безперервні імітаційні моделі зазвичай подаються у вигляді різницево-диференційних рівнянь, які описують взаємодію між різними елементами системи. Дискретні моделі описують системи, поведінка яких змінюється тільки в певні моменти часу. Типовим прикладом такої моделі є черга, що являє собою систему, зміни в якій відбуваються лише тоді, коли клієнт надходить у чергу або залишає систему після обслуговування. Це означає, що в будь-якій дискретній імітаційній моделі є дві головних події, за якими необхідно досліджувати систему. В імітаційній моделі події, пов'язані із прибуттям, визначаються часом між надходженнями клієнтів, а події, пов'язані з їхнім доглядом, - часом обслуговування.

Випадковість в імітаційних моделях виникає тоді, коли інтервал часу t між однорідними подіями є випадковим. Відомий ряд методів одержання послідовних випадкових значень $t=t_1$, t_2, …, що мають заданий розподіл імовірностей $f(x)$. Розглянемо два з них: метод зворотних функцій і метод згорток.

Обидва методи засновані на використанні незалежних однаково розподілених випадкових чисел, що мають рівномірний розподіл на інтервалі [0,1]. Метод зворотних функцій використовується для безперервних розподілів, наприклад для експонентного або рівномірного. Метод згорток використовується в складніших ситуаціях, наприклад, при генеруванні випадкових чисел, що мають нормальний розподіл або розподіл Пуассона.

Метод зворотних функцій вимагає виконання наступних дій. Спочатку генерується випадкове число R з інтервалу [0,1], потім обчислюється шукане випадкове число $x=F^{-1}(R)$, де F^{-1}– функція, зворотна до функції розподілу $F(x)=P\{X<x\}$ (рис. 9.1). Метод заснований на тому, що якщо функцію розподілу $F(x)$ розглядати як випадкову величину, то вона розподілена рівномірно на інтервалі [0,1].

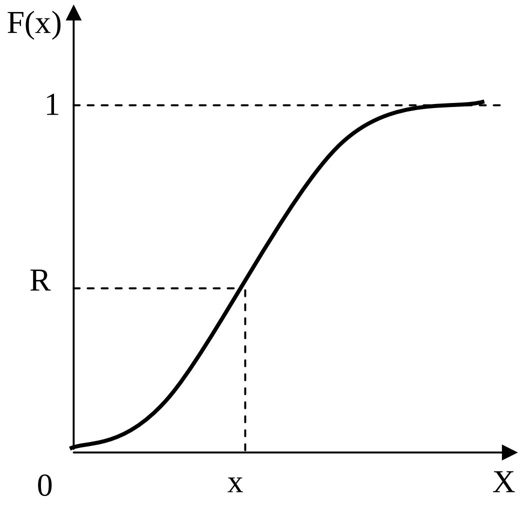

Рисунок 9.1 – Метод зворотних функцій

Розглянемо приклад. Нехай час появи клієнтів розподілений за експонентним законом з параметром λ=4. Функція розподілу має вигляд $F(t)=1-e^{-\lambda t}=1-e^{-4t}$. Урахуємо, що $F(t)=R$, одержимо $t=\frac{1}{\lambda}\ln(1-R)$. Оскільки R - випадкове число з інтервалу [0,1] і (1-R) також випадкове число з того самого інтервалу, можна замінити (1-R) на R. Нехай R=0,9, тоді одержимо одно конкретне значення інтервалу часу між клієнтами $t=\frac{1}{4}\ln(1-0{,}9)=0{,}577$.

Значення R мають обиратися випадково з інтервалу [0,1] і підпорядковуватися рівномірному розподілу.

Основна ідея методу згорток полягає в тому, щоб виразити шукану випадкову величину у вигляді суми інших випадкових величин, для яких легко отримати реалізації випадкових значень. Для одержання значень, що відповідають нормальному розподілу з математичним сподіванням m і стандартним відхиленням σ використовують центральну граничну теорему. Нагадаємо, що сутність її збігається до того, що сума n однаково розподілених випадкових величин прагне до нормального розподілу при нескінченому збільшенні n. Нехай $x=R_1+R_2+\ldots+R_n$, де R_1, R_2, …, R_n – випадкові числа, рівномірно розподілені в інтервалі [0,1]. Відповідно до центральної граничної теореми випадкова величина x є асимптотично нормальною величиною з середнім $n/2$ і дисперсією $n/12$. Тоді випадкова величина X з математичним сподіванням m і стандартним відхиленням σ визначається за формулою

$$X=m+\sigma\left(\frac{x-\frac{n}{2}}{\sqrt{\frac{n}{12}}}\right).$$

У практичних розрахунках для зручності зазвичай приймають n=12, тоді $X=m+\sigma(x-6)$.

Імітаційне моделювання являє собою статистичний експеримент. Його результати мають ґрунтуватися на відповідних статистичних перевірках з використанням, наприклад, довірчих інтервалів і методів перевірки гіпотез. Для цього спостереження повинні задовольняти наступним вимогам: мати стаціонарні розподіли, тобто такі, що не змінюються під час проведення експерименту, підпорядковуватися нормальному розподілу та бути незалежними.

Характер імітаційних обчислень стимулює створення спеціалізованих мов програмування. У цей час на ринку програмних продуктів для моделювання домінують комерційні пакети Arena, AweSim і GPSS/H, які мають розвинений інтерфейс, що спрощує процес створення імітаційних моделей.

9.4 Прийняття рішень в умовах невизначеності та ризику

У теорії прийняття рішень використовують процедури вибору найкращої з кількох можливих альтернатив. Наскільки правильним буде вибір, залежить від якості даних, використовуваних під час опису ситуації, у якій приймається рішення. Із цього погляду процес прийняття рішень може належати до однієї з трьох можливих умов. Прийняття рішень в умовах визначеності, коли дані відомі точно; прийняття рішень в умовах ризику, коли дані можна описати за допомогою імовірнісних розподілів; прийняття рішень в умовах невизначеності, коли даним не можна приписати вагові коефіцієнти, які зумовлювали би ступінь їхньої значущості в процесі прийняття рішень.

Якщо рішення приймається в умовах ризику, вагові коефіцієнти альтернативних рішень описують імовірнісними розподілами. Для ухвалення рішення в цьому випадку використовують критерій сподіваного значення, відповідно до якого альтернативні рішення порівнюють з погляду максимізації сподіваного прибутку або мінімізації сподіваних витрат. При цьому припускається, що прибуток (витрати), пов'язаний з кожним альтернативним рішенням, є випадковою величиною. Розглянемо ситуацію, пов'язану з ухваленням рішення при наявності кінцевого числа альтернатив і точних значень матриці доходів.

Необхідно вкласти на фондовій біржі 10 тис.дол. в акції однієї з двох компаній *A* або *B*. Акції компанії *A* є ризикованими, але можуть принести 50% прибутку від суми інвестиції протягом року. Якщо умови фондової біржі будуть несприятливі, сума інвестиції може знецінитися на 20%. Компанія *B* забезпечує безпеку інвестицій з 15% прибутку в умовах підвищення котирувань на біржі й тільки 5% в умовах зниження котирувань. Аналітичні публікації з імовірністю 60% прогнозують підвищення котирувань. У яку компанію краще вкласти гроші? Складемо таблицю 9.1.

Таблиця 9.1 – Вихідні дані

Альтернативні рішення	**Прибуток від інвестиції 10 тис. дол.**	
	При підвищенні котирувань	**При зниженні котирувань**
Акції компанії *A*	5000	-2000
Акції компанії *B*	1500	500
Імовірність	0,6	0,4

Визначимо сподіваний прибуток. Для акцій компанії *A*

$$5000 \times 0{,}6 + (-2000) \times 0{,}4 = 2200 \text{ дол.}$$

Для акцій компанії *B*

$$1500 \times 0{,}6 + 500 \times 0{,}4 = 1100 \text{ дол.}$$

Таким чином, рішенням, заснованим на цих обчисленнях, є покупка акцій компанії *A*.

У цьому випадку підвищення й зниження котирувань на біржі є станами природи. У загальному випадку задача прийняття рішень може включати n станів природи й m альтернатив. Якщо p_j – імовірність j-го стану природи, а a_{ij}- платіж, пов'язаний з ухваленням рішення, то очікуваний платіж для рішення i обчислюється в такий спосіб

$$M[v_i] = a_{i1}p_1 + a_{i2}p_2 + ... + a_{in}p_n, \quad где \quad \sum p_j = 1.$$

Найкращим рішенням буде те, що відповідає $M[v_i]^* = \max\{M[v_i]\}$ або $M[v_i]^* = \min\{M[v_i]\}$ залежно від того, чи є платіж у задачі доходом або збитком.

Критерій сподіваного значення має дві модифікації, перша з яких полягає у визначенні апостеріорних імовірностей на підставі експерименту над досліджуваною системою, а друга – у визначенні функції корисності.

Імовірності, які використовуються для формулювання критерію сподіваного значення, визначаються, як правило, на підставі попередньо накопиченої інформації. Іноді виявляється можливим перерахувати ці імовірності за допомогою поточної інформації, отриманої на підставі вибіркових (або експериментальних) даних. Одержувані при цьому імовірності називають апостеріорними (або беєсовськими) на відміну від апріорних, отриманих з вихідної інформації.

Функцію корисності $U(x)$ використовують, коли важливіше не реальна величина платежів, а скоріше їхня корисність, зумовлена ставленням особи, що приймає рішення (ОПР), до ризику. Нехай ОПР має шанс 50 на 50, що інвестиція в 20 тис. дол. або принесе прибуток в 40 тис. дол. або буде цілком загублена. Різні індивідууми проявляють різне ставлення до ризику, тобто вони виявляють стосовно ризику різну корисність. У розглянутій ситуації найкращий платіж дорівнює 40 тис. дол., а найгірший становить -20 тис. дол. Установимо довільну шкалу корисності U – від 0 до 100. 0 відповідає корисності -20, а 100 – 40, тобто $U(-20)=0$; $U(40) =100$. Для визначення загального вигляду функції корисності визначимо корисність у точках між -20 і 40.

Якщо ставлення до ризику ОПР байдуже, то функція корисності є прямою, що з'єднує точки (-20, 0) і (40, 100), як показано на рисунку 9.2. У цьому випадку як реальні гроші, так і їхня корисність дають співпадаючі рішення. Нехай ОПР X не прихильне до ризику, тоді для нього при зміні в 10 тис. дол. вправо та уліво від 0 збільшення прибутку змінює корисність на величини ab і bc відповідно, причому, $ab < bc$. До таких самих змін ОПР Y ставиться інакше, у цьому випадку $de>ef$. Визначимо корисність, що відповідає проміжним значенням -10; 0; 10; 20. Функція корисності обчислюється за формулою

$$U(x)=p(-20)+(1-p)U(40).$$

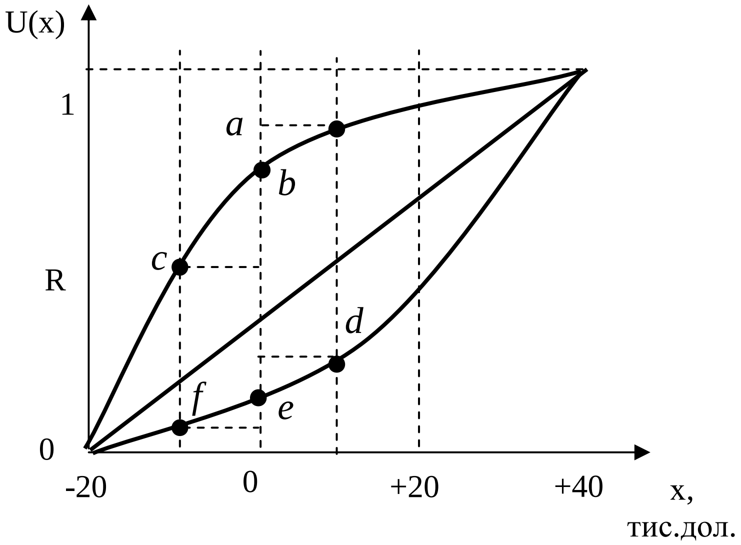

Рисунок 9.2 – Функція корисності

Для визначення значення $U(x)$ просять ОПР повідомити свою перевагу між гарантованою наявною сумою x і можливістю з імовірністю p програти 20 тис. дол. або з імовірністю ($1-p$) виграти 40 тис. дол. Під перевагою тут розуміють вибір такої ймовірності p, за якої пропоновані варіанти однаково привабливі. Наприклад, при x=20 тис. дол. ОПР повідомляє, що p=0,8, тоді $U(20)$ =100-100*0,8=20. Аналогічно знаходять ряд точок для визначення форми функції корисності, а потім визначають шукану криву за допомогою регресійного аналізу або лінійної інтерполяції.

Прийняття рішень в умовах невизначеності, як і в умовах ризику, вимагає визначення альтернативних дій, яким відповідають «платежі», що залежать від станів природи. Запишемо матрицю платежів у задачі прийняття рішень з m можливими діями та n станами природи

$$\begin{array}{c|cccc|}
 & \omega_1 & \omega_2 & \ldots & \omega_n \\
a_1 & v(a_1, \omega_{1)} & v(a_1, \omega_{2)} & \ldots & v(a_1, \omega_n) \\
a_2 & v(a_2, \omega_{1)} & v(a_2, \omega_{2)} & \ldots & v(a_1, \omega_n) \\
\ldots & \ldots & \ldots & \ldots & \ldots \\
a_m & v(a_m, \omega_{1)} & v(a_m, \omega_{2)} & \ldots & v(a_m, \omega_n)
\end{array}$$

Елемент a_i являє i-й можливий розв'язок, а елемент ω_j - j-й стан природи. Плата або дохід, пов'язаний з розв'язком a_i і станом ω_j дорівнює $v(a_i, \omega_j)$.

Відмінність між прийняттям рішень в умовах ризику та невизначеності полягає в тому, що в умовах невизначеності імовірнісний розподіл, що відповідає станам ω_j, невідомий. Ця обставина зумовила розвиток спеціальних критеріїв для аналізу ситуації, пов'язаної з прийняттям рішень: критерію Лапласа, мінімаксного критерію, критерію Севіджа та критерію Гурвіца. Ці критерії відрізняються за ступенем консерватизму особи, що приймає рішення.

Критерій Лапласа спирається на принцип недостатньої підстави, який говорить, що оскільки розподіл імовірностей станів $P\{\omega_j\}$ невідомий, немає підстав вважати їх різними. Отже робиться припущення, що імовірності всіх станів природи дорівнюють одна одній, тобто $P\{\omega_1\}= P\{\omega_2\}=\ldots=P\{\omega_n\}=1/n$. Якщо при цьому $v(a_i, \omega_j)$ є прибутком, то найкращим рішенням є те, що забезпечує

$$\max_{ai}\left\{\frac{1}{n}\sum_{j=1}^{n} v(a_i,\omega_j)\right\}.$$

Якщо величина $v(a_i, \omega_j)$ являє собою видатки, то максимум замінюється на мінімум.

Максимінний (мінімаксний) критерій заснований на консервативному обережному ставленні особи, що приймає рішення, і збігається до вибору найкращої альтернативи з найгірших. Якщо величина $v(a_i, \omega_j)$ є прибутком, то відповідно до максимінного критерію за оптимальне обирається рішення, що забезпечує

$$\max_{ai}\left\{\min_{\omega j} v(a_i,\omega_j)\right\}.$$

Якщо величина $v(a_i, \omega_j)$ являє собою втрати, використовується мінімаксний критерій

$$\min_{ai}\left\{\max_{\omega j} v(a_i,\omega_j)\right\}.$$

Критерій Севіджа знижує консерватизм мінімаксного (максимінного) критерію шляхом заміни матриці платежів $v(a_i, \omega_j)$ матрицею втрат $r(a_i, \omega_j)$, що визначається в такий спосіб

$$r(a_i,\omega_j)=\begin{cases} \max\limits_{ak}\left\{v(a_k,\omega_j)\right\}-v(a_i,\omega_j), & \text{якщо } v-\text{дохід},\\ v(a_{ш},\omega_j)-\min\limits_{ak}\left\{v(a_k,\omega_j)\right\}, & \text{якщо } v-\text{втрати}.\end{cases}$$

Розглянемо приклад. Є матриця платежів

	ω_1	ω_2
a_1	11000	90
a_2	10000	10000

Максимум рядка 1 становить 11000, а рядка 2 – 10000 (мінімакс). Застосування мінімаксного критерію призводить до того, що рішення a_2 з фіксованими втратами 10000 є кращим. Проте можна вибрати й a_1, тому що в цьому випадку існує можливість втратити лише 90, якщо реалізується стан ω_2 при потенційному виграші 11000.

Визначимо, який результат вийде, якщо в мінімаксному критерії замість матриці платежів $v(a_i, \omega_j)$ скористатися матрицею втрат $r(a_i, \omega_j)$.

$$\begin{array}{c|cc|}
 & \omega_1 & \omega_2 \\
\hline
a_1 & 1000 & 0 \\
a_2 & 0 & 9910
\end{array}$$

Максимум рядка 1 становить 1000 (мінімакс), а рядка 2 – 9910. Як бачимо, мінімаксний критерій, застосований до матриці втрат, призводить до вибору рішення a_1.

Критерій Гурвіца охоплює ряд різних підходів до прийняття рішень – від найоптимістичнішого до найпесимістичнішого. Нехай $0 \le \alpha \le 1$ і величини $v(a_i, \omega_j)$ є доходами. Тоді рішенню, обраному за критерієм Гурвіца, відповідає

$$\max_{ai}\left\{\alpha \max_{\omega j} v(a_i, \omega_j) + (1-\alpha)\min_{\omega j} v(a_i, \omega_j)\right\}.$$

Параметр α - показник оптимізму. Якщо $\alpha=0$, то критерій Гурвіца стає еквівалентним застосуванню мінімаксного критерію. Якщо $\alpha=1$, критерій Гурвіца стає занадто оптимістичним, тому що розраховує на найкраще з найкращих рішень. Можна конкретизувати ступінь оптимізму вибором величини α з інтервалу [0, 1]. Найбільш розумним представляється вибір $\alpha=0{,}5$.

Якщо величини $v(a_i, \omega_j)$ є втратами, то критерій приймає наступний вигляд

$$\min_{ai}\left\{\alpha \min_{\omega j} v(a_i, \omega_j) + (1-\alpha)\max_{\omega j} v(a_i, \omega_j)\right\}.$$

Контрольні запитання

1. Поясніть суть задач стохастичного програмування.
2. Яка стохастична задача називається одноетапною? Багатоетапною? Що розуміють під етапами?
3. Дайте загальну характеристику методів стохастичного програмування.
4. Які класи задач стохастичного програмування прийнято розрізняти?
5. Поясніть розходження між стратегічною й статистичною іграми.
6. Для чого використовують метод імітаційного моделювання? У чому відмінність безперервних і дискретних імітаційних моделей?
7. Для чого призначений метод зворотних функцій? Поясніть суть методу.
8. Для чого призначений метод згорток? Поясніть суть методу.
9. Яка відмінність між прийняттям рішень в умовах визначеності, в умовах невизначеності й в умовах ризику?
10. У яких випадках для ухвалення рішення використовують критерій сподіваного значення?

ТЕМА 10
ЕЛЕМЕНТИ ТЕОРІЇ ІГОР

10.1 Основні поняття теорії ігор

При вивченні процесів прийняття рішень кількома суб'єктами, інтереси яких можуть не збігатися, виникають задачі з багатьма цільовими функціями (критеріями). Область математики, що вивчає дані проблеми, одержала назву теорії ігор. Задачі теорії ігор належать до області прийняття рішень в умовах невизначеності, а їхня специфіка полягає в тому, що, як правило, мається на увазі невизначеність, яка виникає в результаті дій двох або більше «розумних» супротивників, здатних оптимізувати свою поведінку за рахунок інших. Такі ситуації є типовими в практичній діяльності менеджерів, маркетологів та інших фахівців, що приймають рішення в умовах гострої конкуренції, неповноти інформації та ін.

Одним з основних питань у задачах з колективним вибором рішень є питання про визначення оптимальності, тобто питання, які рішення можна визнавати найкращими в ситуації оптимізації за кількома критеріями, що відбиває різні інтереси. Багато методів вирішення проблем теорії ігор ґрунтуються на зведенні їх до задач математичного програмування. Теорія ігор бере початок від робіт Е. Бореля (1921 р.), а принциповим етапом у її становленні як самостійного наукового напрямку стала монографія Дж. Неймана, що вийшла в 1944 р.

Термін **гра** застосовується для позначення сукупності правил і угод, якими керуються суб'єкти, поведінку яких ми вивчаємо. Кожний такий суб'єкт k, де $k=\overline{1,K}$, або **гравець**, характеризується наявністю індивідуальної системи цільових настанов і **стратегій** $s_1^k, s_2^k, ..., s_{m_k}^k$, тобто можливих варіантів дій у грі.

Досить розповсюджений спосіб математичного опису гри заснований на завданні функцій $f_k(s_{i1}^1, s_{i2}^2, ..., s_{ik}^k, ..., s_{iK}^K)$, кожна з яких визначає результат (платіж, виграш), одержуваний k-м гравцем залежно від набору стратегій $S=(s_{i1}^1, s_{i2}^2, ..., s_{ik}^k, ..., s_{iK}^K)$, застосованого всіма учасниками гри. Функції f_k також називають **функціями виграшу**, або **платіжними функціями**. У тому випадку, якщо для будь-яких S

$$\sum_{k=1}^{K} f_k(s) = 0,$$

гра називається грою з нульовою сумою. Гру з двома учасниками й нульовою сумою називають антагоністичною. Антагоністичні ігри, тобто ігри, у яких виграш одного учасника дорівнює програшу іншого, у зв'язку з відносно простою постановкою задачі є найбільш вивченим розділом теорії ігор. Проте зміст теорії ігор не вичерпується ними. У класифікації ігрових моделей виділяють ігри з кінцевими й нескінченними наборами стратегій у гравців, виділяють ігри за можливими кількостями ходів в учасників. Також ігри поділяють на некооперативні й кооперативні, тобто ті, в яких функції виграшу учасників залежать від утворених ними коаліцій. Окрім цього ігри можна розрізняти за обсягом інформації, наявної в гравців щодо минулих ходів. У цьому зв'язку вони поділяються на ігри з повною й неповною інформацією.

Розглянемо докладніше антагоністичні ігри та їхні основні властивості. Зручним способом завдання гри двох учасників з нульовою сумою є платіжна матриця. Звідси випливає ще одна їхня назва - матричні ігри. Кожний елемент платіжної матриці a_{ij} містить числове значення виграшу гравця *I* (програшу гравця *II*), якщо перший застосовує стратегію *i*, а другий - стратегію *j*. Терміни виграш і програш потрібно розуміти в широкому змісті, тому що вони можуть приймати від'ємні значення й означати протилежне. Нетривіальність задачі, насамперед, полягає в тім, що кожний з гравців робить свій вибір, не знаючи про вибір іншого. Це істотно ускладнює процес оптимізації обираної стратегії.

Класичним прикладом антагоністичної гри є гра з двома учасниками, що загадують незалежно один від одного числа. Передбачається, що якщо їхня сума виявляється парною, то виграш, що дорівнює 1, дістається першому гравцеві, а якщо непарною, - то другому. Поклавши, що для обох гравців загадування непарного числа є першою стратегією, а парного - другою, можемо записати платіжну матрицю даної гри:

	н/п	п
н/п	1	-1
п	-1	1

(10.1)

Рядки матриці (10.1) відповідають стратегіям гравця *I*, стовпці - стратегіям гравця *II*, а її елементи - результатам першого гравця. Також з визначення гри випливає, що елементи даної матриці, що узяті зі зворотним знаком, відповідають виграшам другого гравця.

Більш складну й змістовнішу платіжну матрицю можна отримати, якщо декілька модифікувати запропоновану гру. Припустимо, що обоє учасники мають право загадувати числа від 1 до 4, що становить їхні відповідні стратегії. У випадку, якщо результат додавання задуманих чисел буде парним, то другий гравець виплачує першому суму, що вийшла, а якщо непарним, то перший - другому. Запишемо платіжну матрицю для такої гри:

$$A=\begin{array}{c|cccc|c} & 2 & -3 & 4 & -5 & 1 \\ & -3 & 4 & -5 & 6 & 2 \\ & 4 & -5 & 6 & -7 & 3 \\ & -5 & 6 & -7 & 8 & 4 \\ \multicolumn{1}{c}{} & \multicolumn{1}{c}{1} & 2 & 3 & \multicolumn{1}{c}{4} & \end{array} \qquad (10.2)$$

Найважливішим у теорії ігор є питання про оптимальність розв'язку (вибору стратегії) для кожного з гравців. Проаналізуємо з цього погляду певну матричну гру, для якої задана платіжна матриця $A = \left\| a_{ij} \right\|_{m \times n}$. При виборі гравцем

I стратегії *i* його гарантований дохід незалежно від дій гравця *II* складе $\min_j a_{ij}$. Оскільки він може вибирати *i* самостійно, то доцільно цей вибір зробити таким, щоб він при будь-якій стратегії супротивника максимізував величину гарантованого доходу, тобто забезпечував одержання $\max_i \left(\min_j a_{ij} \right)$. Такий принцип вибору стратегії одержав назву **принцип максиміна**, а розмір гарантованого виграшу – **нижньої ціни гри**. З іншого боку, аналогічні міркування можна провести і з приводу дій другого гравця. Його найбільший програш при виборі стратегії *j* складе $\max_i a_{ij}$, і, отже, йому треба вибирати стратегію так, щоб мінімізувати величину програшу при будь-яких діях суперника, тобто забезпечити $\min_j \left(\max_i a_{ij} \right)$. у цьому полягає суть **принципу мінімакса**, розмір програшу називається **верхньою ціною гри**.

Можна довести справедливість наступного співвідношення:

$$\max_i \min_j a_{ij} \le \min_j \max_i a_{ij} . \tag{10.3}$$

Проте, очевидний інтерес представляє ситуація, за якої значення виграшу (платежу), одержуваного гравцем *I* при виборі ним максимінної стратегії, дорівнює платежу (програшу) *П*-го гравця при мінімаксній стратегії

$$\max_i \min_j a_{ij} = \min_j \max_i a_{ij} . \tag{10.4}$$

У цьому випадку говорять, що гра має **сідлову точку**. Збіг значень гарантованих виграшів гравців при максимінній і мінімаксній стратегії означає можливість досягнення в грі певного оптимального (рівноважного) стану, від якого невигідно відхилятися жодному з учасників. Поняття «оптимальність» тут означає, що жоден розумний, тобто обережний, гравець не прагне змінити свою стратегію, тому що його супротивник зможе вибрати таку стратегію, що дасть гірший для першого результат. Стратегії *i** і *j**, що утворюють сідлову точку, називаються **оптимальними**, а значення $v = a_{i^* j^*}$ називають **ціною гри**. Трійка (i*,j*,v) уважається **розв'язком матричної гри із сідловою точкою**.

Неважко помітити, що не усяка гра має сідлову точку. Зокрема, як гра (10.1), так і гра (10.2) сідлової точки не мають. Прикладом гри, що має сідлову точку, є гра з платіжною матрицею:

$$A = \begin{vmatrix} 7 & 10 & 4 & 1 \\ 6 & 8 & 5^* & 17 \\ 8 & -3 & 2 & 11 \end{vmatrix} \tag{10.5}$$

У даній матриці мінімальні (гарантовані) виграші першого гравця за рядками дорівнюють 1, 5 і (-3). Отже, його максимінному вибору відповідатиме стратегія 2, що гарантує виграш 5. Для другого гравця максимальні програші за стовпцями матриці складатимуть 8, 10, 5, 17, тому має сенс зупинитися на стратегії 3, за якої він програє тільки 5. Отже, друга стратегія першого гравця й третя стратегія другого утворюють сідлову точку зі значенням 5, тобто гра з матрицею (10.5) має розв’язок (2; 3; 5).

10.2 Змішані стратегії. Основна теорема теорії ігор

Подальший розвиток теорії матричних ігор ґрунтується на дослідженні гри як певного повторюваного процесу. Дійсно, навряд чи можна дати змістовні рекомендації з такого питання, як треба поводитися учасникам однократно проведеної гри, що не має сідлової точки. У випадку ж її багаторазових повторів природною і плідною уявляється ідея рандомізації вибору стратегій гравцями, тобто внесення до процесу вибору елементу випадковості. Дійсно, систематичне відхилення, наприклад, гравця *I* від максимінної стратегії з метою збільшення виграшу може зафіксувати другий гравець і покарати. У той самий час абсолютно хаотичний вибір стратегій не принесе в середньому найкращого результату.

Змішаною стратегією гравця *I* у грі з матрицею $A = \|a_{ij}\|_{m\times n}$ називається впорядкований набір дійсних чисел x_i, $i = \overline{1,m}$, що задовольняють умовам

$$x_i \geq 0, \quad i = \overline{1,m}; \quad \sum_{i=1}^{m} x_i = 1. \tag{10.6}$$

Числа x_i інтерпретуються як імовірності застосування гравцем *I* стратегій 1,2, …, m, які, на відміну від змішаних, називають чистими стратегіями.

Аналогічно вводиться поняття змішаних стратегій гравця *II*, які визначаються як набір чисел y_j, $j = \overline{1,n}$, що задовольняють умовам

$$y_j \geq 0, \quad j = \overline{1,n}; \quad \sum_{j=1}^{n} y_j = 1. \tag{10.7}$$

Тоді, якщо гравець *I* застосовує змішану стратегію $x=(x_1, x_2, ..., x_m)$, а гравець *II* змішану стратегію $y = (y_1, y_2, …, y_n)$, математичне сподівання виграшу гравця *I* (програшу гравця *II*) визначається співвідношенням

$$F(x, y) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij} x_i y_j \,. \tag{10.8}$$

Надалі через X будемо позначати множину припустимих змішаних стратегій гравця *I*, зумовлену умовою (10.6), а через Y - зумовлену умовою (10.7) множину припустимих змішаних стратегій гравця *II*.

До пошуку розв’язку гри в змішаних стратегіях можна застосувати критерії максиміна-мінімакса. Відповідно до них гравець *I* обиратиме свою змішану стратегію $x=(x_1, x_2, ..., x_m)$ таким чином, щоб максимізувати найменший середній виграш:

$$\max_{x\in X}\left[\min_{y\in Y} \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij} x_i y_j\right], \tag{10.9}$$

який дорівнює

$$\max_{x\in X}\left[\min\left\{\sum_{i=1}^{m}a_{i1}x_i,\sum_{i=1}^{m}a_{i2}x_i,...,\sum_{i=1}^{m}a_{in}x_i\right\}\right], \qquad (10.10)$$

а гравець *II* - свою змішану стратегію так, щоб мінімізувати найбільший середній програш:

$$\min_{y\in Y}\left[\max_{x\in X}\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}x_iy_j\right], \qquad (10.11)$$

що дорівнює

$$\min_{y\in Y}\left[\max\left\{\sum_{j=1}^{n}a_{1j}y_j,\sum_{j=1}^{n}a_{2j}y_j,...,\sum_{j=1}^{n}a_{mj}y_j\right\}\right]. \qquad (10.12)$$

За аналогією з (10.3) для будь-яких $x\in X$ і $y\in Y$ справедливою є нерівність

$$\max_{x\in X}\left[\min_{y\in Y}\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}x_iy_j\right]\le\min_{y\in Y}\left[\max_{x\in X}\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}x_iy_j\right]. \qquad (10.13)$$

Стратегії $x^*\in X$ і $y^*\in Y$ називають оптимальними змішаними стратегіями, якщо для будь-яких $x\in X$ і $y\in Y$ справедливою є рівність

$$F(x^*,y^*)=\max_{x\in X}\left[\min_{y\in Y}\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}x_iy_j\right]=\min_{y\in Y}\left[\max_{x\in X}\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}x_iy_j\right]. \qquad (10.14)$$

$v = F(x^*,y^*)$ називають ціною гри, і якщо x^* і y^* існують, то говорять, що гра має розв’язок у змішаних стратегіях (x^*, y^*, v).

Справедливою є фундаментальна теорема Дж. Неймана, (приведемо без доказу).

Теорема 10.1 (основна теорема матричних ігор). Будь-яка матрична гра має розв’язок у змішаних стратегіях.

Значення та нетривіальність теореми (10.1) зумовлені насамперед тим, що у загальному випадку матричні ігри в чистих стратегіях розв’язку не мають.

10.3 Зведення гри двох гравців до задачі лінійного програмування

Задача, розв'язувана першим гравцем, (10.10) була сформульована як максимізація найменшої з сум

$$\sum_{i=1}^{m}a_{ij}x_i\ ,$$

але якщо визначити деяке x_{m+1}, для якого виконується

$$x_{m+1}\le\sum_{i=1}^{m}a_{ij}x_i\ , \qquad (10.15)$$

то її можна звести до задачі лінійного програмування:

знайти $F(x)\Rightarrow\max$ (10.16)

при обмеженнях

$$\sum_{i=1}^{m}a_{ij}x_i\ge x_{m+1},\quad j=\overline{1,n};\quad \sum_{i=1}^{m}x_i=1;\quad x_i\ge0;\quad i=\overline{1,m}. \qquad (10.17)$$

Провівши аналогічні міркування, приходимо до того, що задача мінімізації найбільшого сподіваного програшу, розв'язувана гравцем *II* (10.12), збігається до задачі лінійного програмування

$$F(x) \Rightarrow \min, \qquad (10.18)$$

$$\sum_{j=1}^{n} a_{ij} y_j \le y_{n+1}, \quad i = \overline{1,m}; \quad \sum_{j=1}^{n} y_j = 1; \quad y_j \ge 0; \quad j = \overline{1,n}. \qquad (10.19)$$

Таким чином, ми одержуємо можливість застосовувати всі можливості апарату лінійного програмування для пошуку оптимальних стратегій обох гравців.

Досить легко перевірити, що задачі (10.16)-(10.17) і (10.18)-(10.19) утворюють двоїсту пару. Тут у певному змісті ми повернулися до взаємозв'язку між наявністю розв’язку в певної оптимізаційної задачі та існуванням сідлової точки у відповідної функції Лагранжа. У цьому випадку аналогічний зв'язок простежується між сідловою точкою гри й розв’язком пари задач оптимізації.

Графічний метод вирішення гри. Слід зазначити, що застосування для розв’язання задач (10.16)-(10.17), (10.18)-(10.19) стандартних алгоритмів лінійного програмування далеко не завжди є раціональним. Окрім цього існують інші методи, які ґрунтуються на використанні специфіки даних задач. Зупинимося на дуже простому класичному способі пошуку оптимальних змішаних стратегій у матричних іграх, де один з учасників має тільки дві стратегії (це так звані $2 \times n$ і $m \times 2$ ігри).

Для визначеності покладемо, що гравець *I* має можливість вибирати між двома стратегіями з імовірностями x_1 і x_2=l-x_1, тоді його очікувані виграші, що відповідають чистим стратегіям гравця *II*, приймуть вигляд

$$a_{11}x_1 + a_{21}(1-x_1),\ a_{12}x_1 + a_{23}(1-x_1), \ldots, a_{1n}x_1 + a_{2n}(1-x_1)$$

або

$$(a_{11}-a_{21})x_1 + a_{21},\ (a_{12}-a_{22})x_1 + a_{22}, \ldots, (a_{1n}-a_{2n})x_1 + a_{2n},$$

тобто очікувані виграші можна подати у вигляді графіків лінійних функцій, що залежать від змінної x_1 (рис. 10.1, де передбачається, що гравець *II* має три стратегії).

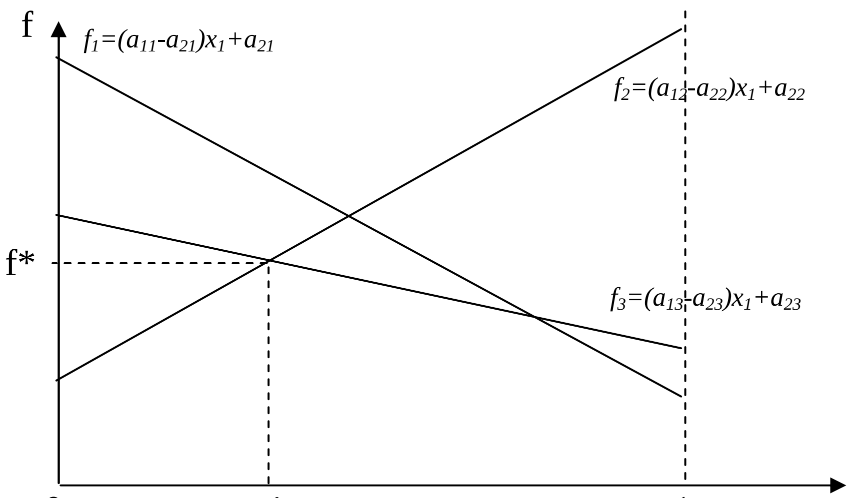

Рисунок 10.1 - Графічне вирішення гри

Лінії, зображені на рисунку 10.1, задають залежності середнього виграшу гравця *I* від значення ймовірності x_1, з якою він обирає свою першу стратегію, для випадків, коли його супротивник обирає першу, другу або третю чисту стратегію. Тоді значенням мінімального гарантованого доходу першого гравця відповідає нижня огинаюча всіх трьох прямих. Відповідно до принципу максиміну, оптимальному вибору гравця *I* відповідатиме найвища точка, що лежить на даній огинаючій, відзначена на рисунку як (x_1*, *f**). Знаючи її, можна визначити оптимальну змішану стратегію першого гравця $x^*=(x_1^*, 1\text{-}x_2^*)$ і ціну гри, що дорівнює *f**.

Виходячи з відношення подвійності, яким пов'язані задачі обох гравців, за оптимальною стратегією першого учасника *x** однозначно визначається оптимальна стратегія його супротивника *y**. Оскільки *y* є результатом розв'язання задачі лінійного програмування, то він має всі властивості припустимого базисного плану, тобто у випадку $2\times n$ гри має не більше за дві ненульових компоненти й не менш за (n-2) нульових. Номера ненульових елементів *y** визначаються номерами ліній, перетинання яких визначило оптимальну стратегію першого гравця. Дійсно, гравець *II* знає оптимальну стратегію суперника, і застосування ним стратегій, що відповідають прямим, які проходять вище точки (x_1*, *f**), тільки збільшило б його програш. У розглянутому прикладі це лінії f_2 і f_3, і, отже, у своїй оптимальній стратегії другий гравець має з ненульовими ймовірностями застосовувати другу й третю чисті стратегії (y_2 >0, y_3 >0). На основі цього, а також з огляду на умову нормування

$$\sum_{j=1}^{n} y_j = 1,$$

можемо виразити: $y_3=1-y_2$, тоді оптимальне значення y^*_2 можна знайти з умови

$$a_{11}\times 0 + a_{21}y_2^* + a_{31}(1-y_2^*) = a_{12}\times 0 + a_{22}y_2^* + a_{32}(1-y_2^*),$$

або

$$(a_{21}-a_{31})y_2^* + a_{31} = (a_{22}-a_{32})y_{21}^* + a_{32}.$$

У результаті одержуємо оптимальну стратегію гравця *II* $y^*=(0, y^*_2, y^*_3)$.

Очевидно, що пошук розв'язку в грі $m\times 2$ здійснюється аналогічно: будують графіки очікуваного програшу гравця *II*, знаходять їхню верхню огинаючу й т.д.

Графічний спосіб у силу обмеженості кола задач, до яких він може бути застосований, має скоріше теоретичне, ніж практичне значення. Проте він добре ілюструє змістовну сторону процесу пошуку розв'язку в грі.

Контрольні запитання

1. Коротко сформулюйте предмет теорії ігор як наукової дисципліни.
2. Який зміст вкладається в поняття «гра»?
3. Для опису яких економічних ситуацій може бути застосований апарат теорії ігор?
4. Яка гра називається антагоністичною?
5. Чим однозначно визначаються матричні ігри?

6. У чому полягають принципи максиміна та мінімакса?
7. При яких умовах можна говорити про те, що гра має сідлову точку?
8. Наведіть приклади ігор, які мають сідлову точку та у яких вона відсутня.
9. Які підходи існують до визначення оптимальних стратегій?
10. Що називають «ціною гри»?
11. Дайте визначення поняттю «змішана стратегія».
12. Сформулюйте основну теорему матричних ігор.

СПИСОК ДЖЕРЕЛ

1. Вітлінський В. В. Математичне програмування: [навч. посіб для студ. вищих навч. закл. економ. спец.] / В. В. Вітлінський, С. І. Наконечний, Т. О.Терещенко. - К: КНЕУ, 2001. - 380 с.

2. Кузнецов Ю. Н. Математическое программирование: учебник / Ю. Н. Кузнецов, В. А. Кузубов, А. В. Волощенко. - М.: Высш.школа, 1980. - 240 с.

3. Акулич И. Л. Математическое програмирование в примерах и задачах : учеб. пособ. / И. Л. Акулич .- М.: Высш. школа, 1986. - 244 с.

4. Таха Х. А. Введение в исследование операций / Х. А. Таха - М.: Изд. дом «Вильямс», 2005. - 912 с.

5. Исследование операций в экономике : Уч. пособие для вузов / Н. Ш. Кремер, Б. А. Путко, И. М. Тришин, М. Н. Фридман./ Под ред. проф. Н. Ш. Кремера. - М.: ЮНИТИ, 2003. - 407 с.

6. Акоф Р. Основы исследования операций / Р. Акоф, М. Сасиени. - М.: Мир, 1971. - 320 с.

Навчальне видання

ВОРОНКОВ Олексій Олександрович

КОНСПЕКТ ЛЕКЦІЙ
з курсу

ОПТИМІЗАЦІЙНІ МЕТОДИ І МОДЕЛІ

(для студентів заочної форми навчання освітньо-кваліфікаційного рівня «бакалавр» напряму підготовки 6.030504 - Економіка підприємства)

Відповідальний за випуск *Т. А. Пушкар*

За авторською редакцією

Комп’ютерне верстання *І. В. Волосожарова*

План 2015, поз. 185 Л

Підп. до друку 12.11.2015 Формат 60x84/16
Друк на ризографі Ум. друк. арк. 6,5
Тираж 50 пр. Зам. №

Видавець і виготовлювач:
Харківський національний університет
міського господарства імені О. М. Бекетова,
вул. Революції, 12, Харків, 61002
Електронна адреса: rectorat@kname.edu.ua
Свідоцтво суб’єкта видавничої справи:
ДК № 4705 від 28.03.2014 р.